

# Entropy and Source Coding for Integer-Dimensional Singular Random Variables

Günther Koliander, Georg Pichler, *Student Member, IEEE*,  
Erwin Riegler, and Franz Hlawatsch, *Fellow, IEEE*

**Abstract**—Entropy and differential entropy are important quantities in information theory. A tractable extension to singular random variables—which are neither discrete nor continuous—has not been available so far. Here, we present such an extension for the practically relevant class of integer-dimensional singular random variables. The proposed entropy definition contains the entropy of discrete random variables and the differential entropy of continuous random variables as special cases. We show that it transforms in a natural manner under Lipschitz functions, and that it is invariant under unitary transformations. We define joint entropy and conditional entropy for integer-dimensional singular random variables, and we show that the proposed entropy conveys useful expressions of the mutual information. As first applications of our entropy definition, we present a result on the minimal expected codeword length of quantized integer-dimensional singular sources and a Shannon lower bound for integer-dimensional singular sources.

**Index Terms**—Information entropy, rate distortion theory, Shannon lower bound, singular random variables, source coding.

## I. INTRODUCTION

### A. Background and Motivation

Entropy is one of the fundamental concepts in information theory. The classical definition of entropy for discrete random variables and its interpretation as information content go back to Shannon [1] and were analyzed thoroughly from axiomatic [2] and operational [1] viewpoints. A similar definition for continuous random variables, differential entropy, was also introduced by Shannon [1], but its interpretation as information content is controversial [3]. Nonetheless, information-theoretic derivations involving undisputed quantities like Kullback-Leibler divergence or mutual information between continuous random variables can often be simplified using differential entropy. Furthermore, in rate-distortion theory, a lower bound on the rate-distortion function known as the Shannon lower bound can be calculated using differential entropy [4, Sec. 4.6].

This paper was presented in part at the IEEE International Symposium on Information Theory (ISIT), Honolulu, HA, July 2014.

G. Koliander, G. Pichler, and F. Hlawatsch are with the Institute of Telecommunications, TU Wien, 1040 Vienna, Austria (e-mail: {gkoliander, gpichler, fhlawatsch}@nt.tuwien.ac.at).

E. Riegler is with the Department of Information Technology and Electrical Engineering, ETH Zurich, CH 8092 Zurich, Switzerland (e-mail: eriegler@nari.ee.ethz.ch).

This work was supported by the WWTF under grants ICT10-066 (NO-WIRE) and ICT12-054 (TINCOIN), and by the FWF under grant P27370-N30.

Copyright (c) 2016 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

Finally, differential entropy arises in asymptotic expansions of the entropy of ever finer quantizations of a continuous random variable [3, Sec. IV]. Hence, although the interpretation of differential entropy is disputed, its operational relevance renders it a useful quantity.

The concepts of entropy and differential entropy thus simplify the understanding and information-theoretic treatment of discrete and continuous random variables. However, these two kinds of random variables do not cover all interesting information-theoretic problems. In fact, a number of information-theoretic problems involving *singular* random variables, which are neither discrete nor continuous, have been described recently:

- For the vector interference channel, a singular input distribution has to be used to fully utilize the available degrees of freedom [5].
- In a probabilistic formulation of analog compression, the underlying source distribution is singular [6].
- In block-fading channel models, two different kinds of singular distributions arise: the optimal input distribution is singular in some settings [7, Ch. 6], and the noiseless output distribution is singular except for special cases [8].

Thus, a suitable generalization of (differential) entropy to singular random variables has the potential to simplify theoretical work in these areas and to provide valuable insights.

Another field where singular random variables appear is source coding. In many high-dimensional problems, deterministic dependencies reduce the intrinsic dimension of a source. Thus, the random variable describing the source cannot be continuous but often is not discrete either. A basic example is a random variable  $\mathbf{x} = (x_1 \ x_2)^T \in \mathbb{R}^2$  supported on the unit circle, i.e., exhibiting the deterministic dependence  $x_1^2 + x_2^2 = 1$ . Although  $\mathbf{x}$  is defined on  $\mathbb{R}^2$  and both components  $x_1, x_2$  are continuous random variables,  $\mathbf{x}$  itself is intrinsically only one-dimensional. The differential entropy of  $\mathbf{x}$  is not defined and, in fact, classical information theory does not provide a rigorous definition of entropy for this random variable. Another, less trivial, example of a singular random variable is a rank-one random matrix of the form  $\mathbf{X} = \mathbf{z}\mathbf{z}^T$ , where  $\mathbf{z}$  is a continuous random vector.

The case of arbitrary probability distributions is very hard to handle, and due to its generality even the mere definition of a meaningful entropy seems impossible. Two existing approaches to defining (differential) entropy for more general distributions are based on quantizations of the random variable in question. Usually, the entropy of these discretizations converges to infinity and, thus, a normalization has to be

employed to obtain a useful result. In [9], this approach is adopted for very specific quantizations of a random variable. Unfortunately, this does not always result in a well-defined entropy and sometimes even fails for continuous random variables of finite differential entropy [9, pp. 197f]. Moreover, the quantization process seems difficult to deal with analytically and no theory was built based on this definition of entropy.<sup>1</sup> A similar approach is to consider arbitrary quantizations that are constrained by some measure of fineness to enable a limit operation. In [3] and [10],  $\varepsilon$ -entropy is introduced as the minimal entropy of all quantizations using sets of diameter less than  $\varepsilon$ . However, to specify a diameter, a distortion function has to be defined. Since all basic information-theoretic quantities (e.g., mutual information or Kullback-Leibler divergence) do not depend on a specific distortion function, it is hardly possible to embed  $\varepsilon$ -entropy into a general information-theoretic framework. Furthermore, once again the quantization process seems difficult to handle analytically.

Since the aforementioned approaches do not provide a satisfactory generalization of (differential) entropy, we follow a different approach, which is also motivated by ever finer quantizations of the random variable. However, in our approach, the order of the two steps “taking the limit of quantizations” and “calculating the entropy as the expectation of the logarithm of a probability (mass) function” is changed. More precisely, we first consider the probability mass functions of quantizations and take a normalized limit. (In the special case of a continuous random variable, this results in the probability density function due to Lebesgue’s differentiation theorem.) Then we take the expectation of the logarithm of the resulting density function. Due to fundamental results in geometric measure theory, this approach can result in a well-defined entropy only for integer-dimensional distributions, since otherwise the density function does not exist [11, Th. 3.1]. In fact, the existence of the density function implies that the random variable is distributed according to a *rectifiable measure* [11, Th. 1.1]. Thus, the distributions considered in the present paper are rectifiable distributions on Euclidean space. Although this is still far from the generality of arbitrary probability distributions, it covers numerous interesting cases—including all the examples mentioned above—and gives valuable insights.

The density function of rectifiable measures can also be defined as a certain Radon-Nikodym derivative. A generalized (differential) entropy based on a Radon-Nikodym derivative with respect to a “measure of the observer’s interest” was considered in [12]. Our entropy is consistent with this approach, and at a certain point we will use a result on quantization problems established in [12]. However, because in our setting a concrete measure is considered, the results we obtain go beyond the basic properties derived in [12] for general measures.

## B. Contributions

We provide a generalization of the classical concepts of entropy and differential entropy to integer-dimensional random

variables. Our entropy satisfies several well-known properties of differential entropy: it is invariant under unitary transformations, transforms as expected under Lipschitz mappings, and can be extended to joint and conditional entropy. We show that the entropy of discrete random variables and the differential entropy of continuous random variables are special cases of our entropy definition. For joint entropy, we prove a chain rule which takes the geometry of the support set into account. Furthermore, we discuss why in certain cases our entropy definition may violate the classical result that conditioning does not increase (differential) entropy. We provide expressions of the mutual information between integer-dimensional random variables in terms of our entropy. We also show that an asymptotic equipartition property analogous to [13, Sec. 8.2] holds for our entropy, but with the Lebesgue measure replaced by the Hausdorff measure of appropriate dimension.

In our proofs, we exercise care to detail all assumptions and to obtain mathematically rigorous statements. Thus, although many of our results might seem obvious to the cursory reader because of their similarity to well-known results for (differential) entropy, we emphasize that they are not simply replicas or straightforward adaptations of known results. This becomes evident, e.g., for the chain rule (see Theorem 41 in Section VI-C), which might be expected to have the same form as the chain rule for differential entropy. However, already a simple example will show that the geometry of the support set may lead to an additional term, which is not present in the special case of continuous random variables.

As a first application of the proposed entropy, we derive a result on the minimal expected binary codeword length of quantized integer-dimensional singular sources. More specifically, we show that our entropy characterizes the rate at which an arbitrarily fine quantization of an integer-dimensional singular source can be compressed. Another application is a lower bound on the rate-distortion function of an integer-dimensional singular source that resembles the Shannon lower bound for discrete [4, Sec. 4.3] and continuous [4, Sec. 4.6] random variables. For the specific case of a singular source that is uniformly distributed on the unit circle, we demonstrate that our bound is within 0.2 nat of the true rate-distortion function.

## C. Notation

Sets are denoted by calligraphic letters (e.g.,  $\mathcal{A}$ ). The complement of a set  $\mathcal{A}$  is denoted  $\mathcal{A}^c$ . Sets of sets are denoted by fraktur letters (e.g.,  $\mathfrak{M}$ ). The set of natural numbers  $\{1, 2, \dots\}$  is denoted as  $\mathbb{N}$ . The open ball with center  $\mathbf{x} \in \mathbb{R}^M$  and radius  $r > 0$  is denoted by  $\mathcal{B}_r(\mathbf{x})$ , i.e.,  $\mathcal{B}_r(\mathbf{x}) \triangleq \{\mathbf{y} \in \mathbb{R}^M : \|\mathbf{y} - \mathbf{x}\| < r\}$ . The symbol  $\omega(M)$  denotes the volume of the  $M$ -dimensional unit ball, i.e.,  $\omega(M) = \pi^{M/2}/\Gamma(1 + M/2)$  where  $\Gamma$  is the Gamma function. Boldface uppercase and lowercase letters denote matrices and vectors, respectively. The  $m \times m$  identity matrix is denoted by  $\mathbf{I}_m$ . Sans serif letters denote random quantities, e.g.,  $\mathbf{x}$  is a random vector and  $x$  is a random scalar. The superscript  $\top$  stands for transposition. For  $x \in \mathbb{R}$ ,  $\lfloor x \rfloor \triangleq \max\{m \in \mathbb{Z} : m \leq x\}$  and for  $\mathbf{x} \in \mathbb{R}^M$ ,  $\lfloor \mathbf{x} \rfloor \triangleq (\lfloor x_1 \rfloor \cdots \lfloor x_M \rfloor)^\top$ . Similarly,  $\lceil x \rceil \triangleq \min\{m \in \mathbb{Z} : m \geq x\}$ . We write  $\mathbb{E}_{\mathbf{x}}[\cdot]$  for the expectation operator with respect to the random variable

<sup>1</sup>This entropy should not be confused with the *information dimension* defined in the same paper [9], which is indeed a very useful and widely used tool.

$\mathbf{x}$ .  $\Pr\{\mathbf{x} \in \mathcal{A}\}$  denotes the probability that  $\mathbf{x} \in \mathcal{A}$ . For  $\mathbf{x} \in \mathbb{R}^{M_1}$  and  $\mathbf{y} \in \mathbb{R}^{M_2}$ , we denote by  $\mathbf{p}_{\mathbf{x}}: \mathbb{R}^{M_1+M_2} \rightarrow \mathbb{R}^{M_1}$ ,  $\mathbf{p}_{\mathbf{x}}(\mathbf{x}, \mathbf{y}) = \mathbf{x}$ , the projection of  $\mathbb{R}^{M_1+M_2}$  to the first  $M_1$  components. Similarly,  $\mathbf{p}_{\mathbf{y}}: \mathbb{R}^{M_1+M_2} \rightarrow \mathbb{R}^{M_2}$ ,  $\mathbf{p}_{\mathbf{y}}(\mathbf{x}, \mathbf{y}) = \mathbf{y}$ , denotes the projection of  $\mathbb{R}^{M_1+M_2}$  to the last  $M_2$  components. The generalized Jacobian determinant of a Lipschitz function<sup>2</sup>  $\phi$  is written as  $J_{\phi}$ . For a function  $\phi$  with domain  $\mathcal{D}$  and a subset  $\tilde{\mathcal{D}} \subseteq \mathcal{D}$ , we denote by  $\phi|_{\tilde{\mathcal{D}}}$  the restriction of  $\phi$  to the domain  $\tilde{\mathcal{D}}$ .  $\mathcal{H}^m$  denotes the  $m$ -dimensional Hausdorff measure.<sup>3</sup>  $\mathcal{L}^M$  denotes the  $M$ -dimensional Lebesgue measure and  $\mathfrak{B}_M$  denotes the Borel  $\sigma$ -algebra on  $\mathbb{R}^M$ . For a measure  $\mu$  and a  $\mu$ -measurable function  $f$ , the induced measure is given as  $\mu f^{-1}(\mathcal{A}) \triangleq \mu(f^{-1}(\mathcal{A}))$ . For two measures  $\mu$  and  $\nu$  on the same measurable space, we indicate by  $\mu \ll \nu$  that  $\mu$  is absolutely continuous with respect to  $\nu$  (i.e., for any measurable set  $\mathcal{A}$ ,  $\nu(\mathcal{A}) = 0$  implies  $\mu(\mathcal{A}) = 0$ ). For a measure  $\mu$  and a measurable set  $\mathcal{E}$ , the measure  $\mu|_{\mathcal{E}}$  is the restriction of  $\mu$  to  $\mathcal{E}$ , i.e.,  $\mu|_{\mathcal{E}}(\mathcal{A}) = \mu(\mathcal{A} \cap \mathcal{E})$ . The logarithm to the base  $e$  is denoted  $\log$  and the logarithm to the base 2 is denoted  $\text{ld}$ . In certain equations, we reference an equation number on top of the equality sign in order to indicate that the equality holds due to some previous equation: e.g.,  $\stackrel{(42)}{=}$  indicates that the equality holds due to eq. (42).

#### D. Organization of the Paper

The rest of this paper is organized as follows. In Section II, we review the established definitions of entropy and describe the intuitive idea behind our entropy definition. Rectifiable sets, measures, and random variables are introduced in Section III as the basic setting for integer-dimensional distributions. In Section IV, we develop the theory of “lower-dimensional entropy”: we define entropy for integer-dimensional random variables, prove a transformation property and invariance under unitary transformations, demonstrate connections to classical entropy and differential entropy, and provide examples by calculating the entropy of random variables supported on the unit circle in  $\mathbb{R}^2$  and of positive semidefinite rank-one random matrices. In Sections V and VI, we introduce and discuss joint entropy and conditional entropy, respectively. Relations of our entropy to the mutual information between integer-dimensional random variables are demonstrated in Section VII. In Section VIII, we prove an asymptotic equipartition property for our entropy. In Section IX, we present a result on the minimal expected binary codeword length of quantized integer-dimensional sources. In Section X, we derive a Shannon lower bound for integer-dimensional singular sources and evaluate it for a source that is uniformly distributed on the unit circle.

## II. PREVIOUS WORK AND MOTIVATION

We first recall the definitions of entropy for discrete random variables [13, Ch. 2] and differential entropy for continuous

random variables [13, Ch. 8]. Let  $\mathbf{x}$  be a discrete random variable with probability mass function  $p_{\mathbf{x}}(\mathbf{x}_i) = \Pr\{\mathbf{x} = \mathbf{x}_i\}$ ,  $i \in \mathcal{I}$ , where  $\mathcal{I}$  is the finite or countably infinite set indexing all possible realizations  $\mathbf{x}_i$  of  $\mathbf{x}$ . The entropy of  $\mathbf{x}$  is

$$H(\mathbf{x}) \triangleq -\mathbb{E}_{\mathbf{x}}[\log p_{\mathbf{x}}(\mathbf{x})] = -\sum_{i \in \mathcal{I}} p_{\mathbf{x}}(\mathbf{x}_i) \log p_{\mathbf{x}}(\mathbf{x}_i). \quad (1)$$

For a continuous random variable  $\mathbf{x}$  on  $\mathbb{R}^M$  with probability density function  $f_{\mathbf{x}}$ , the differential entropy is

$$h(\mathbf{x}) \triangleq -\mathbb{E}_{\mathbf{x}}[\log f_{\mathbf{x}}(\mathbf{x})] = -\int_{\mathbb{R}^M} f_{\mathbf{x}}(\mathbf{x}) \log f_{\mathbf{x}}(\mathbf{x}) d\mathcal{L}^M(\mathbf{x}). \quad (2)$$

We note that  $h(\mathbf{x})$  may be  $\pm\infty$  or undefined.

#### A. Entropy of Dimension $d(\mathbf{x})$ and $\varepsilon$ -Entropy

There exist two previously proposed generalizations of (differential) entropy to a larger set of probability distributions. The first generalization is based on quantizations of the random variable to ever finer cubes [9]. More specifically, for a (possibly singular) random variable  $\mathbf{x} \in \mathbb{R}^M$ , the *Rényi information dimension* of  $\mathbf{x}$  is

$$d(\mathbf{x}) \triangleq \lim_{n \rightarrow \infty} \frac{H\left(\frac{\lfloor n\mathbf{x} \rfloor}{n}\right)}{\log n} \quad (3)$$

and the *entropy of dimension*  $d(\mathbf{x})$  of  $\mathbf{x}$  is defined as

$$h_{d(\mathbf{x})}^{\mathbf{R}}(\mathbf{x}) \triangleq \lim_{n \rightarrow \infty} \left( H\left(\frac{\lfloor n\mathbf{x} \rfloor}{n}\right) - d(\mathbf{x}) \log n \right) \quad (4)$$

provided the limits in (3) and (4) exist.

This definition of entropy of dimension  $d(\mathbf{x})$  corresponds to the following procedure:

- 1) Quantize  $\mathbf{x}$  using the cubes  $\prod_{i=1}^M \left[\frac{k_i}{n}, \frac{k_i+1}{n}\right)$ , with  $k_i \in \mathbb{Z}$ , i.e., consider the discrete random variable with probabilities  $p_{\mathbf{k}} = \Pr\{\mathbf{x} \in \prod_{i=1}^M \left[\frac{k_i}{n}, \frac{k_i+1}{n}\right)\}$ .
- 2) Calculate the entropy of the quantized random variable, i.e., the negative expectation of the logarithm of the probability mass function  $p_{\mathbf{k}}$ .
- 3) Subtract the correction term  $d(\mathbf{x}) \log n$  to account for the dimension of the random variable  $\mathbf{x}$ .
- 4) Take the limit  $n \rightarrow \infty$ .

Although this approach seems reasonable, there are several issues. First, the definition of  $h_{d(\mathbf{x})}^{\mathbf{R}}(\mathbf{x})$  seems to be difficult to handle analytically, and connections to major information-theoretic concepts such as mutual information are not available. Furthermore, the quantization used is just one of many possible—we might, e.g., consider a shifted version of the set of cubes  $\prod_{i=1}^M \left[\frac{k_i}{n}, \frac{k_i+1}{n}\right)$ , which, for singular distributions, may result in a different value of the resulting entropy.

An approach that overcomes the latter issue is the concept of  $\varepsilon$ -entropy [3], [10]. The definition of  $\varepsilon$ -entropy does not use a specific quantization but takes the infimum of the entropy over all possible (countable) quantizations under a constraint on the diameter of the quantization sets. This is motivated by data compression: the quantization should be such that an error of maximally  $\varepsilon$  is made (thus, the quantization sets have maximal diameter  $\varepsilon$ ) and at the same time the minimal possible number of bits should be used to encode the data (thus, the entropy is

<sup>2</sup>By Rademacher’s theorem [14, Th. 2.14], a Lipschitz function is differentiable almost everywhere and, thus, the Jacobian determinant is well defined almost everywhere.

<sup>3</sup>Readers unfamiliar with this concept may think of it as a measure of an  $m$ -dimensional area in a higher-dimensional space (e.g., surfaces in  $\mathbb{R}^3$ ). An introduction and definition can be found in [14, Sec. 2.8].

minimized over all possible quantizations). More specifically, for a random variable  $\mathbf{x} \in \mathbb{R}^M$ , let  $\mathfrak{P}_\varepsilon$  denote the set of all countable partitions of  $\mathbb{R}^M$  into mutually disjoint, measurable sets of diameter at most  $\varepsilon$ . Furthermore, for a partition  $\Omega = \{\mathcal{A}_i : i \in \mathbb{N}\} \in \mathfrak{P}_\varepsilon$ , the quantization  $[\mathbf{x}]_\Omega \in \mathbb{N}$  is the discrete random variable defined by  $p_i = \Pr\{[\mathbf{x}]_\Omega = i\} = \Pr\{\mathbf{x} \in \mathcal{A}_i\}$  for  $i \in \mathbb{N}$ . Then the  $\varepsilon$ -entropy of  $\mathbf{x}$  is defined as

$$H_\varepsilon(\mathbf{x}) \triangleq \inf_{\Omega \in \mathfrak{P}_\varepsilon} H([\mathbf{x}]_\Omega). \quad (5)$$

Here, a problem is that  $H_\varepsilon(\mathbf{x})$  is only defined for a fixed  $\varepsilon > 0$  and the limit  $\varepsilon \rightarrow 0$  converges to  $\infty$  for nondiscrete distributions. However, as in the case of Rényi information dimension, a correction term can be obtained using the following seemingly new definition of information dimension:

$$d^*(\mathbf{x}) \triangleq \lim_{\varepsilon \rightarrow 0} \frac{H_\varepsilon(\mathbf{x})}{\log \frac{1}{\varepsilon}}.$$

By [15, Prop. 3.3], the definitions of information dimension using Rényi's approach and the  $\varepsilon$ -entropy approach coincide, i.e.,  $d^*(\mathbf{x}) = d(\mathbf{x})$ . This suggests the following new definition of a  $d(\mathbf{x})$ -dimensional entropy.

*Definition 1:* Let  $\mathbf{x} \in \mathbb{R}^M$  be a random variable with existing information dimension  $d(\mathbf{x})$ . Then the *asymptotic  $\varepsilon$ -entropy of dimension  $d(\mathbf{x})$*  is defined as

$$h_{d(\mathbf{x})}^*(\mathbf{x}) \triangleq \lim_{\varepsilon \rightarrow 0} (H_\varepsilon(\mathbf{x}) + d(\mathbf{x}) \log \varepsilon).$$

This definition corresponds to the following procedure:

- 1) Quantize  $\mathbf{x}$  using an entropy-minimizing quantization<sup>4</sup>  $\Omega$  given a diameter constraint  $\varepsilon$ , i.e., consider the discrete random variable  $[\mathbf{x}]_\Omega$  with probabilities  $p_i = \Pr\{[\mathbf{x}]_\Omega = i\} = \Pr\{\mathbf{x} \in \mathcal{A}_i\}$  for  $\mathcal{A}_i \in \Omega$ , where the diameter of each  $\mathcal{A}_i$  is upper bounded by  $\varepsilon$ .
- 2) Calculate the entropy of the quantized random variable  $[\mathbf{x}]_\Omega$ , i.e., the negative expectation of the logarithm of the probability mass function  $p_i$ .
- 3) Add the correction term  $d(\mathbf{x}) \log \varepsilon$  to account for the dimension of the random variable  $\mathbf{x}$ .
- 4) Take the limit  $\varepsilon \rightarrow 0$ .

Although this entropy is more general than the entropy of dimension  $d(\mathbf{x})$  in (4), the fundamental problems persist: we are still restricted to the choice of sets of small diameter (this is of course useful if we consider maximal distance as a measure of distortion but can yield unnecessarily many quantization points for areas of almost zero probability), and the definition still seems to be difficult to handle analytically and lacks connections to established information-theoretic quantities such as mutual information.

### B. An Alternative Approach

Here, we propose a different approach, which is motivated by the definition of differential entropy. The basic idea is to circumvent the quantization step and perform the entropy calculation at the end. Assuming  $\mathbf{x} \in \mathbb{R}^M$ , this results in the following procedure:

- 1) For some  $\mathbf{x} \in \mathbb{R}^M$ , divide the probability  $\Pr\{\mathbf{x} \in \mathcal{B}_\varepsilon(\mathbf{x})\}$  by the correction factor<sup>5</sup>  $\omega(d(\mathbf{x})) \varepsilon^{d(\mathbf{x})}$ . (Recall that  $\omega(d(\mathbf{x}))$  denotes the volume of the  $d(\mathbf{x})$ -dimensional unit ball.)
- 2) Take the limit  $\varepsilon \rightarrow 0$ .
- 3) Calculate the entropy as the negative expectation of the logarithm of the resulting density function.

More specifically, steps 1–2 yield the density function<sup>6</sup>

$$\theta_{\mathbf{x}}(\mathbf{x}) \triangleq \lim_{\varepsilon \rightarrow 0} \frac{\Pr\{\mathbf{x} \in \mathcal{B}_\varepsilon(\mathbf{x})\}}{\omega(d(\mathbf{x})) \varepsilon^{d(\mathbf{x})}} \quad (6)$$

and the entropy in step 3 is thus given by

$$\mathfrak{h}^{d(\mathbf{x})}(\mathbf{x}) \triangleq -\mathbb{E}_{\mathbf{x}}[\log \theta_{\mathbf{x}}(\mathbf{x})]. \quad (7)$$

We will show that this definition of entropy will lead to definitions of joint and conditional entropy, various useful relations, connections to mutual information, an asymptotic equipartition property, and bounds relevant to source coding. However, our definition does have one limitation: as pointed out in [6, Sec. VII-A], the existence of the limit in (6) for almost every  $\mathbf{x} \in \mathbb{R}^M$  is a much stronger assumption than the existence of the Rényi information dimension (3). Loosely speaking, the existence of the limit in (6) requires that the random variable  $\mathbf{x}$  is  $d(\mathbf{x})$ -dimensional almost everywhere whereas the existence of the Rényi information dimension merely requires that the random variable is  $d(\mathbf{x})$ -dimensional “on average.” By Preiss’ Theorem [16, Th. 5.6], convergence in (6) even implies that the probability measure induced by the random variable  $\mathbf{x}$  is rectifiable (see Definition 6 in Section III-B), which means that our definition does not apply to, e.g., self-similar fractal distributions. However, we are not aware of any application or calculation of the  $d(\mathbf{x})$ -dimensional entropy in (4) (or the asymptotic version of  $\varepsilon$ -entropy) for fractal distributions, and it does not seem clear whether the  $d(\mathbf{x})$ -dimensional entropy is well defined in that case (although the information dimension (3) exists).

The rectifiability also implies that the density function  $\theta_{\mathbf{x}}(\mathbf{x})$  is equal to a certain Radon-Nikodym derivative. Based on this equality, the entropy  $\mathfrak{h}^{d(\mathbf{x})}(\mathbf{x})$  defined in (7) and (6) can be interpreted as a *generalized entropy* as defined in [12, eq. (1.5)] by

$$H_\lambda(\mu) \triangleq \begin{cases} -\int_{\mathbb{R}^M} \log \left( \frac{d\mu}{d\lambda}(\mathbf{x}) \right) d\mu(\mathbf{x}) & \text{if } \mu \ll \lambda \\ \infty & \text{else.} \end{cases} \quad (8)$$

Here,  $\lambda$  is a  $\sigma$ -finite measure on  $\mathbb{R}^M$  and  $\mu$  is a probability measure on  $\mathbb{R}^M$ . While  $\mu$  can be chosen as the measure of a given random variable, the generalized entropy (8) provides no intuition on how to choose the measure  $\lambda$ . It is more similar to a divergence between measures and, in particular, reduces to the Kullback-Leibler divergence [17] for a probability measure  $\lambda$ . We will see (cf. Remark 19) that our entropy definition coincides with (8) for the choice  $\lambda = \mathcal{H}^m|_{\mathcal{E}}$ , where  $m$  and

<sup>5</sup>The constant factor  $\omega(d(\mathbf{x}))$  is included to obtain equality with differential entropy in the special case  $d(\mathbf{x}) = M$ . A different factor would result in an additive constant in the entropy definition.

<sup>6</sup>A mathematically rigorous definition will be provided in Section III-B.

<sup>4</sup>We assume for simplicity that an entropy-minimizing quantization exists although in general the infimum in (5) may not be attained.

$\mathcal{E}$  depend on the given random variable. This interpretation will allow us to use basic results from [12] for our entropy definition.

Motivated by the entropy expression in (7), a formal definition of the entropy of an integer-dimensional random variable will be given in Section IV-A, based on the mathematical theory of rectifiable measures discussed next.

### III. RECTIFIABLE RANDOM VARIABLES

As mentioned in Section II-B, the existence of a  $d(\mathbf{x})$ -dimensional density implies that the random variable  $\mathbf{x}$  is rectifiable. In this section, we recall the definitions of rectifiable sets and measures and introduce rectifiable random variables as a straightforward extension. Furthermore, we present some basic properties that will be used in subsequent sections. For the convenience of readers who prefer to skip the mathematical details, we summarize the most important facts in Corollary 12.

#### A. Rectifiable Sets

Our basic geometric objects of interest are rectifiable sets [18, Sec. 3.2.14]. As the definition of rectifiable sets is not consistent in the literature, we provide the definition most convenient for our purpose. We recall that  $\mathcal{H}^m$  denotes the  $m$ -dimensional Hausdorff measure.

**Definition 2** ([14, Def. 2.57]): For  $m \in \mathbb{N}$ , an  $\mathcal{H}^m$ -measurable set  $\mathcal{E} \subseteq \mathbb{R}^M$  (with  $M \geq m$ ) is called  *$m$ -rectifiable*<sup>7</sup> if there exist  $\mathcal{L}^m$ -measurable, bounded sets  $\mathcal{A}_k \subseteq \mathbb{R}^m$  and Lipschitz functions  $f_k: \mathcal{A}_k \rightarrow \mathbb{R}^M$ , both for<sup>8</sup>  $k \in \mathbb{N}$ , such that  $\mathcal{H}^m(\mathcal{E} \setminus \bigcup_{k \in \mathbb{N}} f_k(\mathcal{A}_k)) = 0$ . A set  $\mathcal{E} \subseteq \mathbb{R}^M$  is called 0-rectifiable if it is finite or countably infinite.

**Remark 3:** Hereafter, we will often consider the setting of  $m$ -rectifiable sets in  $\mathbb{R}^M$  and tacitly assume  $m \in \{0, \dots, M\}$ .

Rectifiable sets satisfy the following well-known basic properties.

**Lemma 4:** Let  $\mathcal{E}$  be an  $m$ -rectifiable subset of  $\mathbb{R}^M$ .

- 1) Any  $\mathcal{H}^m$ -measurable subset  $\mathcal{D} \subseteq \mathcal{E}$  is also  $m$ -rectifiable.
- 2) The measure  $\mathcal{H}^m|_{\mathcal{E}}$  is  $\sigma$ -finite.
- 3) Let  $\phi: \mathbb{R}^M \rightarrow \mathbb{R}^N$  with  $N \geq m$  be a Lipschitz function. If  $\phi(\mathcal{E})$  is  $\mathcal{H}^m$ -measurable, then it is  $m$ -rectifiable.
- 4) For  $n > m$ , we have  $\mathcal{H}^n(\mathcal{E}) = 0$ .
- 5) Let  $\mathcal{E}_i$  for  $i \in \mathbb{N}$  be  $m$ -rectifiable sets. Then  $\bigcup_{i \in \mathbb{N}} \mathcal{E}_i$  is  $m$ -rectifiable.
- 6) For  $m \neq 0$ ,  $\mathbb{R}^m$  is  $m$ -rectifiable.

Intuitively, rectifiable sets are lower-dimensional subsets of Euclidean space. Examples include affine subspaces, algebraic varieties, differentiable manifolds, and graphs of Lipschitz functions. As countable unions of rectifiable sets are again rectifiable, further examples are countable unions of any of the aforementioned sets.

**Remark 5:** There are various characterizations of  $m$ -rectifiable sets that provide connections to other mathematical disciplines. For example, an  $\mathcal{H}^m$ -measurable set  $\mathcal{E} \subseteq \mathbb{R}^M$

is  $m$ -rectifiable if and only if there exist  $\mathcal{T}_k \subseteq \mathbb{R}^M$  such that  $\mathcal{E} \subseteq \mathcal{T}_0 \cup \bigcup_{k \in \mathbb{N}} \mathcal{T}_k$ , where  $\mathcal{H}^m(\mathcal{T}_0) = 0$  and each  $\mathcal{T}_k$  is an  $m$ -dimensional, embedded  $C^1$  submanifold of  $\mathbb{R}^M$  [19, Lem. 5.4.2]. Another characterization, based on [18, Cor. 3.2.4], is that  $\mathcal{E} \subseteq \mathbb{R}^M$  is  $m$ -rectifiable if and only if

$$\mathcal{E} \subseteq \mathcal{E}_0 \cup \bigcup_{k \in \mathbb{N}} f_k(\mathcal{A}_k) \quad (9)$$

where  $\mathcal{H}^m(\mathcal{E}_0) = 0$ ,  $\mathcal{A}_k$  are bounded Borel sets, and  $f_k: \mathbb{R}^m \rightarrow \mathbb{R}^M$  are Lipschitz functions that are one-to-one on  $\mathcal{A}_k$ . Due to [20, Th. 15.1], this implies that  $f_k(\mathcal{A}_k)$  are also Borel sets.

#### B. Rectifiable Measures

Loosely speaking, rectifiable measures are measures that are concentrated on a rectifiable set. The most convenient way to define “concentrated on” mathematically is in terms of absolute continuity with respect to a specific Hausdorff measure.

**Definition 6** ([14, Def. 2.59]): A Borel measure  $\mu$  on  $\mathbb{R}^M$  is called  *$m$ -rectifiable* if there exists an  $m$ -rectifiable set  $\mathcal{E} \subseteq \mathbb{R}^M$  such that  $\mu \ll \mathcal{H}^m|_{\mathcal{E}}$ .

For an  $m$ -rectifiable measure  $\mu$ , i.e.,  $\mu \ll \mathcal{H}^m|_{\mathcal{E}}$  for an  $m$ -rectifiable set  $\mathcal{E} \subseteq \mathbb{R}^M$ , we have by Property 2 in Lemma 4 that  $\mathcal{H}^m|_{\mathcal{E}}$  is  $\sigma$ -finite. Thus, by the Radon-Nikodym theorem [14, Th. 1.28], there exists the Radon-Nikodym derivative

$$\theta_\mu^m(\mathbf{x}) \triangleq \frac{d\mu}{d\mathcal{H}^m|_{\mathcal{E}}}(\mathbf{x}) \quad (10)$$

satisfying  $d\mu = \theta_\mu^m d\mathcal{H}^m|_{\mathcal{E}}$ . We will refer to  $\theta_\mu^m(\mathbf{x})$  as the  *$m$ -dimensional Hausdorff density* of  $\mu$ .

**Remark 7:** If  $\mu$  is an  $m$ -rectifiable probability measure, it cannot be  $n$ -rectifiable for  $n \neq m$ . Indeed, suppose that  $\mu$  is both  $m$ -rectifiable and  $n$ -rectifiable where, without loss of generality,  $n > m$ . Then there exists an  $m$ -rectifiable set  $\mathcal{E}$  such that  $\mu \ll \mathcal{H}^m|_{\mathcal{E}}$ , which implies  $\mu(\mathcal{E}^c) = 0$ . There also exists an  $n$ -rectifiable set  $\mathcal{F}$  such that  $\mu \ll \mathcal{H}^n|_{\mathcal{F}}$ . By Property 4 in Lemma 4, the  $m$ -rectifiable set  $\mathcal{E}$  satisfies  $\mathcal{H}^n(\mathcal{E}) = 0$  and, in particular,  $\mathcal{H}^n|_{\mathcal{F}}(\mathcal{E}) = 0$ . Because  $\mu \ll \mathcal{H}^n|_{\mathcal{F}}$ , this implies  $\mu(\mathcal{E}) = 0$ . Hence,  $\mu(\mathbb{R}^M) = \mu(\mathcal{E}^c) + \mu(\mathcal{E}) = 0$ , which is a contradiction to the assumption that  $\mu$  is a probability measure.

To avoid the nuisance of separately considering the case  $\frac{d\mu}{d\mathcal{H}^m|_{\mathcal{E}}} = 0$  in many proofs and to reduce the class of  $m$ -rectifiable sets of interest, we define the following notion of a support of an  $m$ -rectifiable measure.

**Definition 8:** For an  $m$ -rectifiable measure  $\mu$  on  $\mathbb{R}^M$ , an  $m$ -rectifiable set  $\mathcal{E} \subseteq \mathbb{R}^M$  is called a *support* of  $\mu$  if  $\mu \ll \mathcal{H}^m|_{\mathcal{E}}$ ,  $\frac{d\mu}{d\mathcal{H}^m|_{\mathcal{E}}} > 0$   $\mathcal{H}^m|_{\mathcal{E}}$ -almost everywhere, and  $\mathcal{E} = \bigcup_{k \in \mathbb{N}} f_k(\mathcal{A}_k)$  where, for  $k \in \mathbb{N}$ ,  $\mathcal{A}_k$  is a bounded Borel set and  $f_k: \mathbb{R}^m \rightarrow \mathbb{R}^M$  is a Lipschitz function that is one-to-one on  $\mathcal{A}_k$ .

**Lemma 9:** Let  $\mu$  be an  $m$ -rectifiable measure, i.e.,  $\mu \ll \mathcal{H}^m|_{\mathcal{E}}$  for an  $m$ -rectifiable set  $\mathcal{E} \subseteq \mathbb{R}^M$ . Then there exists a support  $\tilde{\mathcal{E}} \subseteq \mathcal{E}$ . Furthermore, the support is unique up to sets of  $\mathcal{H}^m$ -measure zero.

*Proof:* See Appendix A. ■

<sup>7</sup>In [14, Def. 2.57], these sets are called *countably  $\mathcal{H}^m$ -rectifiable*.

<sup>8</sup>This definition also encompasses finite index sets  $k \in \{1, \dots, K\}$ ; it suffices to set  $\mathcal{A}_k = \emptyset$  for  $k > K$ .

*Remark 10:* For  $m$ -rectifiable measures, it is possible to interpret the Hausdorff density  $\theta_\mu^m(\mathbf{x})$  as a measure of “local probability per area.” Indeed, for an  $m$ -rectifiable measure  $\mu$ , i.e.,  $\mu \ll \mathcal{H}^m|_{\mathcal{E}}$  for an  $m$ -rectifiable set  $\mathcal{E}$ , we can write  $\theta_\mu^m(\mathbf{x})$  in (10) as

$$\theta_\mu^m(\mathbf{x}) = \lim_{r \rightarrow 0} \frac{\mu(\mathcal{B}_r(\mathbf{x}))}{\omega(m)r^m} \quad (11)$$

$\mathcal{H}^m|_{\mathcal{E}}$ -almost everywhere (for a proof see [14, Th. 2.83 and eq. (2.42)]). Furthermore, the right-hand side in (11) vanishes for  $\mathcal{H}^m$ -almost all points not in  $\mathcal{E}$ . Note the similarity of (11) with the ad-hoc construction in Section II-B. Indeed, (11) is the mathematically rigorous formulation of (6). This formulation also provides details regarding the probability measures for which it results in a well-defined quantity.

### C. Rectifiable Random Variables

As we are only interested in probability measures and because information theory is often formulated for random variables, we define  $m$ -rectifiable random variables. In what follows, we consider a random variable  $\mathbf{x}: (\Omega, \mathfrak{S}) \rightarrow (\mathbb{R}^M, \mathfrak{B}_M)$  on a probability space  $(\Omega, \mathfrak{S}, \mu)$ , i.e.,  $\Omega$  is a set,  $\mathfrak{S}$  is a  $\sigma$ -algebra on  $\Omega$ , and  $\mu$  is a probability measure on  $(\Omega, \mathfrak{S})$ . The probability measure induced by the random variable  $\mathbf{x}$  is denoted by  $\mu_{\mathbf{x}^{-1}}$ . For  $\mathcal{A} \in \mathfrak{B}_M$ ,  $\mu_{\mathbf{x}^{-1}}(\mathcal{A})$  equals the probability that  $\mathbf{x} \in \mathcal{A}$ , i.e.,

$$\mu_{\mathbf{x}^{-1}}(\mathcal{A}) = \mu(\mathbf{x}^{-1}(\mathcal{A})) = \Pr\{\mathbf{x} \in \mathcal{A}\}. \quad (12)$$

*Definition 11:* A random variable  $\mathbf{x}: \Omega \rightarrow \mathbb{R}^M$  on a probability space  $(\Omega, \mathfrak{S}, \mu)$  is called  $m$ -rectifiable if the induced probability measure  $\mu_{\mathbf{x}^{-1}}$  on  $\mathbb{R}^M$  is  $m$ -rectifiable, i.e., there exists an  $m$ -rectifiable set  $\mathcal{E} \subseteq \mathbb{R}^M$  such that  $\mu_{\mathbf{x}^{-1}} \ll \mathcal{H}^m|_{\mathcal{E}}$ . The  $m$ -dimensional Hausdorff density of an  $m$ -rectifiable random variable  $\mathbf{x}$  is defined as (cf. (10))

$$\theta_{\mathbf{x}}^m(\mathbf{x}) \triangleq \theta_{\mu_{\mathbf{x}^{-1}}}^m(\mathbf{x}) = \frac{d\mu_{\mathbf{x}^{-1}}}{d\mathcal{H}^m|_{\mathcal{E}}}(\mathbf{x}). \quad (13)$$

Furthermore, a support of the measure  $\mu_{\mathbf{x}^{-1}}$  is called a *support of  $\mathbf{x}$* , i.e.,  $\mathcal{E}$  is a support of  $\mathbf{x}$  if  $\mu_{\mathbf{x}^{-1}} \ll \mathcal{H}^m|_{\mathcal{E}}$ ,  $\frac{d\mu_{\mathbf{x}^{-1}}}{d\mathcal{H}^m|_{\mathcal{E}}}(\mathbf{x}) > 0$   $\mathcal{H}^m|_{\mathcal{E}}$ -almost everywhere, and  $\mathcal{E} = \bigcup_{k \in \mathbb{N}} f_k(\mathcal{A}_k)$  where, for  $k \in \mathbb{N}$ ,  $\mathcal{A}_k$  is a bounded Borel set and  $f_k: \mathbb{R}^m \rightarrow \mathbb{R}^M$  is a Lipschitz function that is one-to-one on  $\mathcal{A}_k$ .

Note that due to Remark 7, an  $m$ -rectifiable random variable cannot be  $n$ -rectifiable for  $n \neq m$ .

In the nontrivial case  $m < M$ , the  $m$ -dimensional Hausdorff density  $\theta_{\mathbf{x}}^m(\mathbf{x})$  is not a probability density function in the classical sense and is nonzero only on an  $m$ -dimensional set  $\mathcal{E}$ . Indeed, the random variable  $\mathbf{x}$  will vanish everywhere except on a set of Lebesgue measure zero, and thus a probability density function cannot exist. However, the  $m$ -dimensional Hausdorff measure of the support set does not vanish, and one can think of  $\theta_{\mathbf{x}}^m$  as an  $m$ -dimensional probability density function of the random variable  $\mathbf{x}$  on  $\mathbb{R}^M$ .

Based on our discussion of rectifiable measures in Section III-B, we can find a characterization of  $m$ -rectifiable random variables that resembles well-known properties of continuous random variables. This characterization is stated

in the next corollary. Note, however, that although everything seems to be similar to the continuous case, Hausdorff measures lack substantial properties of the Lebesgue measure, e.g., the product measure is not always again a Hausdorff measure.

*Corollary 12:* Let  $\mathbf{x}$  be an  $m$ -rectifiable random variable on  $\mathbb{R}^M$ , i.e.,  $\mu_{\mathbf{x}^{-1}} \ll \mathcal{H}^m|_{\mathcal{E}}$  for an  $m$ -rectifiable set  $\mathcal{E} \subseteq \mathbb{R}^M$ . Then there exists the  $m$ -dimensional Hausdorff density  $\theta_{\mathbf{x}}^m$ , and the following properties hold:

- 1) The probability  $\Pr\{\mathbf{x} \in \mathcal{A}\}$  for a measurable set  $\mathcal{A} \subseteq \mathbb{R}^M$  can be calculated as the integral of  $\theta_{\mathbf{x}}^m$  over  $\mathcal{A}$  with respect to the  $m$ -dimensional Hausdorff measure restricted to  $\mathcal{E}$ , i.e.,

$$\Pr\{\mathbf{x} \in \mathcal{A}\} = \mu_{\mathbf{x}^{-1}}(\mathcal{A}) = \int_{\mathcal{A}} \theta_{\mathbf{x}}^m(\mathbf{x}) d\mathcal{H}^m|_{\mathcal{E}}(\mathbf{x}). \quad (14)$$

- 2) The expectation of a measurable function  $f: \mathbb{R}^M \rightarrow \mathbb{R}$  with respect to the random variable  $\mathbf{x}$  can be expressed as

$$\mathbb{E}_{\mathbf{x}}[f(\mathbf{x})] = \int_{\mathbb{R}^M} f(\mathbf{x}) \theta_{\mathbf{x}}^m(\mathbf{x}) d\mathcal{H}^m|_{\mathcal{E}}(\mathbf{x}). \quad (15)$$

- 3) The random variable  $\mathbf{x}$  is in  $\mathcal{E}$  with probability one, i.e.,

$$\Pr\{\mathbf{x} \in \mathcal{E}\} = \mu_{\mathbf{x}^{-1}}(\mathcal{E}) = \int_{\mathcal{E}} \theta_{\mathbf{x}}^m(\mathbf{x}) d\mathcal{H}^m|_{\mathcal{E}}(\mathbf{x}) = 1. \quad (16)$$

- 4) There exists a support  $\tilde{\mathcal{E}} \subseteq \mathcal{E}$  of  $\mathbf{x}$ .

The special cases  $m = 0$  and  $m = M$  reduce to well-known concepts.

*Theorem 13:* Let  $\mathbf{x}$  be a random variable on  $\mathbb{R}^M$ . Then:

- 1)  $\mathbf{x}$  is 0-rectifiable if and only if it is a discrete random variable, i.e., there exists a probability mass function  $p_{\mathbf{x}}(\mathbf{x}_i) = \Pr\{\mathbf{x} = \mathbf{x}_i\} > 0$ ,  $i \in \mathcal{I}$ , where  $\mathcal{I}$  is a finite or countably infinite index set indicating all possible realizations  $\mathbf{x}_i$  of  $\mathbf{x}$ . In this case,  $\theta_{\mathbf{x}}^0 = p_{\mathbf{x}}$  and  $\mathcal{E} = \{\mathbf{x}_i : i \in \mathcal{I}\}$  is a support of  $\mathbf{x}$ .
- 2)  $\mathbf{x}$  is  $M$ -rectifiable if and only if it is a continuous random variable, i.e., there exists a probability density function  $f_{\mathbf{x}}$  such that  $\Pr\{\mathbf{x} \in \mathcal{A}\} = \int_{\mathcal{A}} f_{\mathbf{x}}(\mathbf{x}) d\mathcal{L}^M(\mathbf{x})$ . In this case,  $\theta_{\mathbf{x}}^M = f_{\mathbf{x}}$   $\mathcal{L}^M$ -almost everywhere.

*Proof:* See Appendix B. ■

The following theorem introduces a nontrivial class of  $m$ -rectifiable random variables.

*Theorem 14:* Let  $\mathbf{x}$  be a continuous random variable on  $\mathbb{R}^m$ . Furthermore, let  $\phi: \mathbb{R}^m \rightarrow \mathbb{R}^M$  with  $M \geq m$  be a locally Lipschitz mapping whose  $m$ -dimensional Jacobian determinant<sup>9</sup> satisfies  $J_{\phi}(\mathbf{x}) > 0$   $\mathcal{L}^m$ -almost everywhere, and assume that  $\phi(\mathbb{R}^m)$  is  $\mathcal{H}^m$ -measurable. Then  $\mathbf{y} \triangleq \phi(\mathbf{x})$  is an  $m$ -rectifiable random variable on  $\mathbb{R}^M$ .

*Proof:* According to Definition 11, we have to show that  $\mu_{\mathbf{y}^{-1}} \ll \mathcal{H}^m|_{\mathcal{E}}$  for an  $m$ -rectifiable set  $\mathcal{E} \subseteq \mathbb{R}^M$ . By Properties 1, 3, and 6 in Lemma 4, the set  $\mathcal{E} \triangleq \phi(\mathcal{B}_r(\mathbf{0}))$  is  $m$ -rectifiable ( $\phi$  is Lipschitz on  $\mathcal{B}_r(\mathbf{0})$  for all  $r > 0$ ).

<sup>9</sup>The  $m$ -dimensional Jacobian determinant of  $\phi$  is defined as  $J_{\phi}(\mathbf{x}) = \sqrt{\det(D_{\phi}^T(\mathbf{x}) D_{\phi}(\mathbf{x}))}$ , where  $D_{\phi}(\mathbf{x}) \in \mathbb{R}^{M \times m}$  denotes the Jacobian matrix of  $\phi$ , which is guaranteed to exist almost everywhere. Note in particular that  $J_{\phi}(\mathbf{x})$  is nonnegative.

Hence, by Property 5 in Lemma 4, the set  $\mathcal{E} \triangleq \phi(\mathbb{R}^m) = \bigcup_{r \in \mathbb{N}} \phi(\mathcal{B}_r(\mathbf{0}))$  is  $m$ -rectifiable. Thus, it suffices to show that  $\mu_{\mathbf{y}^{-1}} \ll \mathcal{H}^m|_{\phi(\mathbb{R}^m)}$ , i.e., that for any  $\mathcal{H}^m$ -measurable set  $\mathcal{A} \subseteq \mathbb{R}^M$ ,  $\mathcal{H}^m|_{\phi(\mathbb{R}^m)}(\mathcal{A}) = 0$  implies  $\mu_{\mathbf{y}^{-1}}(\mathcal{A}) = 0$ . To this end, assume first that  $\mathcal{H}^m|_{\phi(\mathbb{R}^m)}(\mathcal{A}) = 0$  for a *bounded*  $\mathcal{H}^m$ -measurable set  $\mathcal{A} \subseteq \mathbb{R}^M$ . Let  $f$  denote the probability density function of  $\mathbf{x}$ . By the generalized change of variables formula [14, eq. (2.47)], we have

$$\begin{aligned} & \int_{\phi^{-1}(\mathcal{A})} f(\mathbf{x}) J_\phi(\mathbf{x}) d\mathcal{L}^m(\mathbf{x}) \\ &= \int_{\phi(\phi^{-1}(\mathcal{A}))} \sum_{\mathbf{x} \in \phi^{-1}(\mathcal{A}) \cap \phi^{-1}(\{\mathbf{y}\})} f(\mathbf{x}) d\mathcal{H}^m(\mathbf{y}) \\ &= \int_{\mathcal{A} \cap \phi(\mathbb{R}^m)} \sum_{\mathbf{x} \in \phi^{-1}(\mathcal{A}) \cap \phi^{-1}(\{\mathbf{y}\})} f(\mathbf{x}) d\mathcal{H}^m(\mathbf{y}) \\ &\stackrel{(a)}{=} 0 \end{aligned} \quad (17)$$

where (a) holds because  $\mathcal{H}^m(\mathcal{A} \cap \phi(\mathbb{R}^m)) = 0$ . Because  $J_\phi(\mathbf{x}) > 0$   $\mathcal{L}^m$ -almost everywhere, (17) implies  $f(\mathbf{x}) = 0$   $\mathcal{L}^m$ -almost everywhere on  $\phi^{-1}(\mathcal{A})$ , and hence  $\int_{\phi^{-1}(\mathcal{A})} f(\mathbf{x}) d\mathcal{L}^m(\mathbf{x}) = 0$ . Thus, we have

$$\mu_{\mathbf{y}^{-1}}(\mathcal{A}) = \mu_{\mathbf{x}^{-1}}(\phi^{-1}(\mathcal{A})) = \int_{\phi^{-1}(\mathcal{A})} f(\mathbf{x}) d\mathcal{L}^m(\mathbf{x}) = 0.$$

For an *unbounded*  $\mathcal{H}^m$ -measurable set  $\mathcal{A} \subseteq \mathbb{R}^M$  satisfying  $\mathcal{H}^m|_{\phi(\mathbb{R}^m)}(\mathcal{A}) = 0$ , following the arguments above, we obtain  $\mu_{\mathbf{y}^{-1}}(\mathcal{A} \cap \mathcal{B}_r(\mathbf{0})) = 0$  for the bounded sets  $\mathcal{A} \cap \mathcal{B}_r(\mathbf{0})$ ,  $r \in \mathbb{N}$ . This implies  $\mu_{\mathbf{y}^{-1}}(\mathcal{A}) \leq \sum_{r \in \mathbb{N}} \mu_{\mathbf{y}^{-1}}(\mathcal{A} \cap \mathcal{B}_r(\mathbf{0})) = 0$ . ■

#### D. Example: Distributions on the Unit Circle

As a basic example of 1-rectifiable singular random variables, we consider distributions on the unit circle in  $\mathbb{R}^2$ , i.e., on  $\mathcal{S}_1 \triangleq \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\| = 1\}$ .

*Corollary 15:* Let  $z$  be a continuous random variable on  $\mathbb{R}$ . Then  $\mathbf{x} = (\mathbf{x}_1 \ \mathbf{x}_2)^T \triangleq (\cos z \ \sin z)^T$  is a 1-rectifiable random variable.

*Proof:* The mapping  $\phi: z \mapsto (\cos z \ \sin z)^T$  is Lipschitz and its Jacobian determinant is identically one. Thus, we can directly apply Theorem 14. ■

This toy example is intuitive and illustrates the concept of  $m$ -rectifiable singular random variables in a very simple setup. In a similar way, one can analyze the rectifiability of distributions on various other geometric structures.

#### E. Example: Positive Semidefinite Rank-One Random Matrices

A less obvious example of an  $m$ -rectifiable singular random variable are positive semidefinite rank-one random matrices, i.e., matrices of the form  $\mathbf{X} = \mathbf{z}\mathbf{z}^T \in \mathbb{R}^{m \times m}$ , where  $\mathbf{z}$  is a continuous random variable on  $\mathbb{R}^m$ .

*Corollary 16:* Let  $\mathbf{z}$  be a continuous random variable on  $\mathbb{R}^m$ . Then the random matrix  $\mathbf{X} \triangleq \mathbf{z}\mathbf{z}^T$  is  $m$ -rectifiable on  $\mathbb{R}^{m^2}$ .

*Proof:* The mapping  $\phi: \mathbf{z} \mapsto \mathbf{z}\mathbf{z}^T$  is locally Lipschitz. Thus, in order to apply Theorem 14, it remains to show that

$J_\phi(\mathbf{z}) > 0$   $\mathcal{L}^m$ -almost everywhere. To calculate the Jacobian matrix  $\mathbf{D}_\phi(\mathbf{z})$ , we stack the columns of the matrix  $\mathbf{z}\mathbf{z}^T$  and differentiate the resulting vector with respect to each element  $z_i$ . It is easily seen that the resulting Jacobian matrix is given by

$$\mathbf{D}_\phi(\mathbf{z}) = \begin{pmatrix} \mathbf{z}\mathbf{e}_1^T + z_1\mathbf{I}_m \\ \mathbf{z}\mathbf{e}_2^T + z_2\mathbf{I}_m \\ \vdots \\ \mathbf{z}\mathbf{e}_m^T + z_m\mathbf{I}_m \end{pmatrix} \quad (18)$$

where  $\mathbf{e}_i$  denotes the  $i$ th unit vector. As long as at least one element  $z_i$  is nonzero,  $\mathbf{D}_\phi(\mathbf{z})$  has full rank. Thus,  $J_\phi(\mathbf{z}) > 0$   $\mathcal{L}^m$ -almost everywhere. ■

*Remark 17:* For the case of positive definite random matrices, i.e.,  $\mathbf{X}_m = \sum_{i=1}^m \mathbf{z}_i \mathbf{z}_i^T$  with independent continuous  $\mathbf{z}_i$ , it is easy to see that the measures induced by these random matrices are absolutely continuous with respect to the  $m(m+1)/2$ -dimensional Lebesgue measure on the space of all symmetric matrices. The intermediate case of positive semidefinite rank-deficient random matrices  $\mathbf{X}_n = \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^T$  for  $n \in \{2, \dots, m-1\}$ , where the  $\mathbf{z}_i \in \mathbb{R}^m$ ,  $i \in \{1, \dots, n\}$  are independent continuous random variables, is considerably more involved because the mapping  $(z_1, \dots, z_n) \mapsto \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^T$  has a vanishing Jacobian determinant almost everywhere. We conjecture that  $\mathbf{X}_n$  is  $(mn - n(n-1)/2)$ -rectifiable, conforming to the dimension of the manifold of all positive semidefinite rank- $n$  matrices with  $n$  distinct eigenvalues.

### IV. ENTROPY OF RECTIFIABLE RANDOM VARIABLES

#### A. Definition

The  $m$ -rectifiable random variables introduced in Definition 11 will be the objects considered in our entropy definition. Due to the existence of the  $m$ -dimensional Hausdorff density  $\theta_{\mathbf{x}}^m$  for these random variables (see (11) and (13)), the heuristic approach described in Section II-B (see (6) and (7)) can be made rigorous.

*Definition 18:* Let  $\mathbf{x}$  be an  $m$ -rectifiable random variable on  $\mathbb{R}^M$ . The  $m$ -dimensional entropy of  $\mathbf{x}$  is defined as

$$\mathfrak{h}^m(\mathbf{x}) \triangleq -\mathbb{E}_{\mathbf{x}}[\log \theta_{\mathbf{x}}^m(\mathbf{x})] = -\int_{\mathbb{R}^M} \log \theta_{\mathbf{x}}^m(\mathbf{x}) d\mu_{\mathbf{x}^{-1}}(\mathbf{x}) \quad (19)$$

provided the integral on the right-hand side exists in  $\mathbb{R} \cup \{\pm\infty\}$ .

By (15), we obtain

$$\mathfrak{h}^m(\mathbf{x}) = -\int_{\mathbb{R}^M} \theta_{\mathbf{x}}^m(\mathbf{x}) \log \theta_{\mathbf{x}}^m(\mathbf{x}) d\mathcal{H}^m|_{\mathcal{E}}(\mathbf{x}) \quad (20)$$

$$= -\int_{\mathcal{E}} \theta_{\mathbf{x}}^m(\mathbf{x}) \log \theta_{\mathbf{x}}^m(\mathbf{x}) d\mathcal{H}^m(\mathbf{x}) \quad (21)$$

where  $\mathcal{E} \subseteq \mathbb{R}^M$  is an arbitrary  $m$ -rectifiable set satisfying  $\mu_{\mathbf{x}^{-1}} \ll \mathcal{H}^m|_{\mathcal{E}}$  (in particular,  $\mathcal{E}$  may be a support of  $\mathbf{x}$ ).

*Remark 19:* For a fixed  $m$ -rectifiable measure  $\mu$ , our entropy definition (19) can be interpreted as a generalized entropy (8) with  $\lambda = \mathcal{H}^m|_{\mathcal{E}}$ . This will allow us to use basic results from [12] for our entropy definition. However, our definition changes the measure  $\lambda$  based on the choice of  $\mu$  and thus is not simply a special case of (8).

### B. Transformation Property

One important property of differential entropy is its invariance under unitary transformations. A similar result holds for  $m$ -dimensional entropy. We can even give a more general result for arbitrary one-to-one Lipschitz mappings.

**Theorem 20:** Let  $\mathbf{x}$  be an  $m$ -rectifiable random variable on  $\mathbb{R}^N$  with  $1 \leq m \leq N$ , finite  $m$ -dimensional entropy  $\mathfrak{h}^m(\mathbf{x})$ , support  $\mathcal{E}$ , and  $m$ -dimensional Hausdorff density  $\theta_{\mathbf{x}}^m$ . Furthermore, let  $\phi: \mathbb{R}^N \rightarrow \mathbb{R}^M$  with  $M \geq m$  be a Lipschitz mapping such that<sup>10</sup>  $J_{\phi}^{\mathcal{E}} > 0$   $\mathcal{H}^m|_{\mathcal{E}}$ -almost everywhere,  $\phi(\mathcal{E})$  is  $\mathcal{H}^m$ -measurable, and  $\mathbb{E}_{\mathbf{x}}[\log J_{\phi}^{\mathcal{E}}(\mathbf{x})]$  exists and is finite. If the restriction of  $\phi$  to  $\mathcal{E}$  is one-to-one, then  $\mathbf{y} \triangleq \phi(\mathbf{x})$  is an  $m$ -rectifiable random variable with  $m$ -dimensional Hausdorff density

$$\theta_{\mathbf{y}}^m(\mathbf{y}) = \frac{\theta_{\mathbf{x}}^m(\phi^{-1}(\mathbf{y}))}{J_{\phi}^{\mathcal{E}}(\phi^{-1}(\mathbf{y}))}$$

$\mathcal{H}^m|_{\phi(\mathcal{E})}$ -almost everywhere, and its  $m$ -dimensional entropy is

$$\mathfrak{h}^m(\mathbf{y}) = \mathfrak{h}^m(\mathbf{x}) + \mathbb{E}_{\mathbf{x}}[\log J_{\phi}^{\mathcal{E}}(\mathbf{x})].$$

*Proof:* See Appendix C. ■

**Remark 21:** Theorem 20 shows that for the special case of a unitary transformation  $\phi$  (e.g., a translation),

$$\mathfrak{h}^m(\phi(\mathbf{x})) = \mathfrak{h}^m(\mathbf{x})$$

because  $J_{\phi}^{\mathcal{E}}(\mathbf{x})$  is identically one in that case.

**Remark 22:** In general, no result resembling Theorem 20 holds for Lipschitz functions  $\phi: \mathbb{R}^N \rightarrow \mathbb{R}^M$  that are not one-to-one on  $\mathcal{E}$ . We can argue as in the proof of Theorem 20 and obtain that  $\mathbf{y} = \phi(\mathbf{x})$  is  $m$ -rectifiable and that the  $m$ -dimensional Hausdorff density is

$$\theta_{\mathbf{y}}^m(\mathbf{y}) = \sum_{\mathbf{x} \in \phi^{-1}(\{\mathbf{y}\})} \frac{\theta_{\mathbf{x}}^m(\mathbf{x})}{J_{\phi}^{\mathcal{E}}(\mathbf{x})}$$

$\mathcal{H}^m|_{\phi(\mathcal{E})}$ -almost everywhere. We then obtain for the  $m$ -dimensional entropy

$$\begin{aligned} \mathfrak{h}^m(\mathbf{y}) &= - \int_{\phi(\mathcal{E})} \left( \sum_{\mathbf{x} \in \phi^{-1}(\{\mathbf{y}\})} \frac{\theta_{\mathbf{x}}^m(\mathbf{x})}{J_{\phi}^{\mathcal{E}}(\mathbf{x})} \right) \\ &\quad \times \log \left( \sum_{\mathbf{x} \in \phi^{-1}(\{\mathbf{y}\})} \frac{\theta_{\mathbf{x}}^m(\mathbf{x})}{J_{\phi}^{\mathcal{E}}(\mathbf{x})} \right) d\mathcal{H}^m(\mathbf{y}) \\ &\stackrel{(a)}{=} - \int_{\mathcal{E}} \theta_{\mathbf{x}}^m(\mathbf{x}) \\ &\quad \times \log \left( \sum_{\mathbf{x}' \in \phi^{-1}(\{\phi(\mathbf{x})\})} \frac{\theta_{\mathbf{x}'}^m(\mathbf{x}')}{J_{\phi}^{\mathcal{E}}(\mathbf{x}')} \right) d\mathcal{H}^m(\mathbf{x}) \end{aligned}$$

where (a) holds because of the generalized area formula [14, Th. 2.91]. In most cases, this cannot be easily expressed in terms of a differential entropy due to the sum in the logarithm. However, in the special case of a Jacobian determinant  $J_{\phi}^{\mathcal{E}}$  and a Hausdorff density  $\theta_{\mathbf{x}}^m$  that are symmetric in the sense that  $\theta_{\mathbf{x}}^m(\mathbf{x}')$  and  $J_{\phi}^{\mathcal{E}}(\mathbf{x}')$  are constant on  $\phi^{-1}(\{\phi(\mathbf{x})\})$  for all  $\mathbf{x} \in$

<sup>10</sup>Here  $J_{\phi}^{\mathcal{E}}$  denotes the Jacobian determinant of the tangential differential of  $\phi$  in  $\mathcal{E}$ . For details see [18, Sec. 3.2.16].

$\mathcal{E}$ , the summation reduces to a multiplication by the cardinality of  $\phi^{-1}(\{\phi(\mathbf{x})\})$ .

### C. Relation to Entropy and Differential Entropy

In the special cases  $m = 0$  and  $m = M$ , our entropy definition reduces to classical entropy (1) and differential entropy (2), respectively.

**Theorem 23:** Let  $\mathbf{x}$  be a random variable on  $\mathbb{R}^M$ . If  $\mathbf{x}$  is a 0-rectifiable (i.e., discrete) random variable, then the 0-dimensional entropy of  $\mathbf{x}$  coincides with the classical entropy, i.e.,  $\mathfrak{h}^0(\mathbf{x}) = H(\mathbf{x})$ . If  $\mathbf{x}$  is an  $M$ -rectifiable (i.e., continuous) random variable, then the  $M$ -dimensional entropy of  $\mathbf{x}$  coincides with the differential entropy, i.e.,  $\mathfrak{h}^M(\mathbf{x}) = h(\mathbf{x})$ .

*Proof:* Let  $\mathbf{x}$  be a 0-rectifiable random variable. By Theorem 13,  $\mathbf{x}$  is a discrete random variable with possible realizations  $\mathbf{x}_i$ ,  $i \in \mathcal{I}$ , the 0-dimensional Hausdorff density  $\theta_{\mathbf{x}}^0$  is the probability mass function of  $\mathbf{x}$ , and a support is given by  $\mathcal{E} = \{\mathbf{x}_i : i \in \mathcal{I}\}$ . Thus, (21) yields

$$\begin{aligned} \mathfrak{h}^0(\mathbf{x}) &= - \int_{\mathcal{E}} \theta_{\mathbf{x}}^0(\mathbf{x}) \log \theta_{\mathbf{x}}^0(\mathbf{x}) d\mathcal{H}^0(\mathbf{x}) \\ &\stackrel{(a)}{=} - \sum_{i \in \mathcal{I}} \Pr\{\mathbf{x} = \mathbf{x}_i\} \log \Pr\{\mathbf{x} = \mathbf{x}_i\} \\ &\stackrel{(1)}{=} H(\mathbf{x}) \end{aligned}$$

where (a) holds because  $\mathcal{H}^0$  is the counting measure.

Let  $\mathbf{x}$  be an  $M$ -rectifiable random variable. By Theorem 13,  $\mathbf{x}$  is a continuous random variable and the  $M$ -dimensional Hausdorff density  $\theta_{\mathbf{x}}^M$  is equal to the probability density function  $f_{\mathbf{x}}$ . Thus, (19) yields

$$\mathfrak{h}^M(\mathbf{x}) = -\mathbb{E}_{\mathbf{x}}[\log \theta_{\mathbf{x}}^M(\mathbf{x})] = -\mathbb{E}_{\mathbf{x}}[\log f_{\mathbf{x}}(\mathbf{x})] \stackrel{(2)}{=} h(\mathbf{x}).$$

To get an idea of the  $m$ -dimensional entropy of random variables in between the discrete and continuous cases, we can use Theorem 14 to construct  $m$ -rectifiable random variables. More specifically, we consider a continuous random variable  $\mathbf{x}$  on  $\mathbb{R}^m$  and a one-to-one Lipschitz mapping  $\phi: \mathbb{R}^m \rightarrow \mathbb{R}^M$  ( $M \geq m$ ) whose generalized Jacobian determinant satisfies  $J_{\phi} > 0$   $\mathcal{L}^m$ -almost everywhere. Intuitively, we should see a connection between the differential entropy of  $\mathbf{x}$  and the  $m$ -dimensional entropy of  $\mathbf{y} \triangleq \phi(\mathbf{x})$ . By Theorem 14, the random variable  $\mathbf{y}$  is  $m$ -rectifiable and, because  $\phi$  is one-to-one, we can indeed calculate the  $m$ -dimensional entropy.

**Corollary 24:** Let  $\mathbf{x}$  be a continuous random variable on  $\mathbb{R}^m$  with finite differential entropy  $h(\mathbf{x})$  and probability density function  $f_{\mathbf{x}}$ . Furthermore, let  $\phi: \mathbb{R}^m \rightarrow \mathbb{R}^M$  ( $M \geq m$ ) be a one-to-one Lipschitz mapping such that  $J_{\phi} > 0$   $\mathcal{L}^m$ -almost everywhere and  $\mathbb{E}_{\mathbf{x}}[\log J_{\phi}(\mathbf{x})]$  exists and is finite. Then the  $m$ -dimensional Hausdorff density of the  $m$ -rectifiable random variable  $\mathbf{y} \triangleq \phi(\mathbf{x})$  is

$$\theta_{\mathbf{y}}^m(\mathbf{y}) = \frac{f_{\mathbf{x}}(\phi^{-1}(\mathbf{y}))}{J_{\phi}(\phi^{-1}(\mathbf{y}))}$$

$\mathcal{H}^m|_{\phi(\mathbb{R}^m)}$ -almost everywhere, and the  $m$ -dimensional entropy of  $\mathbf{y}$  is

$$\mathfrak{h}^m(\mathbf{y}) = h(\mathbf{x}) + \mathbb{E}_{\mathbf{x}}[\log J_{\phi}(\mathbf{x})].$$



For the special case of the embedding  $\phi: \mathbb{R}^m \rightarrow \mathbb{R}^M$ ,  $\phi(x_1, \dots, x_m) = (x_1 \dots x_m 0 \dots 0)^T$ , this results in

$$\mathfrak{h}^m(\mathbf{x}_1, \dots, \mathbf{x}_m, 0, \dots, 0) = h(\mathbf{x}). \quad (22)$$

*Proof:* The first part is the special case  $N = m$  and  $\mathcal{E} = \mathbb{R}^m$  of Theorem 20. The result (22) then follows from the fact that, for the considered embedding,  $J_\phi(\mathbf{x})$  is identically 1. ■

#### D. Example: Entropy of Distributions on the Unit Circle

It is now easy to calculate the entropy of the 1-rectifiable singular random variables on the unit circle previously considered in Section III-D. Let  $\mathbf{z}$  be a continuous random variable on  $\mathbb{R}$  with probability density function  $f_z$  supported on  $[0, 2\pi)$ , i.e.,  $f_z(z) = 0$  for  $z \notin [0, 2\pi)$ . By Corollary 24, the 1-dimensional Hausdorff density of the random variable  $\mathbf{x} = \phi(\mathbf{z}) = (\cos z \ \sin z)^T$  is given by (recall that the Jacobian determinant is identically one)

$$\theta_{\mathbf{x}}^1(\mathbf{x}) = f_z(\phi^{-1}(\mathbf{x})) \quad (23)$$

$\mathcal{H}^1|_{\mathcal{S}_1}$ -almost everywhere, and the entropy of  $\mathbf{x}$  is given by

$$\mathfrak{h}^1(\mathbf{x}) = h(\mathbf{z}). \quad (24)$$

Of course, this result for  $\mathfrak{h}^1(\mathbf{x})$  may have been conjectured by heuristic reasoning. Next, we consider a case where heuristic reasoning does not help.

#### E. Example: Entropy of Positive Semidefinite Rank-One Random Matrices

As a more challenging example, we calculate the entropy of a specific type of  $m$ -rectifiable singular random variables, namely, the positive semidefinite rank-one random matrices previously considered in Section III-E.

*Theorem 25:* Let  $\mathbf{z}$  be a continuous random variable on  $\mathbb{R}^m$  with probability density function  $f_z$ , and let  $\bar{\mathbf{z}}$  denote the random variable with probability density function  $f_{\bar{\mathbf{z}}}(\mathbf{z}) = (f_z(\mathbf{z}) + f_z(-\mathbf{z}))/2$ . Then the  $m$ -dimensional entropy of the random matrix  $\mathbf{X} = \mathbf{z}\mathbf{z}^T$  is given by

$$\mathfrak{h}^m(\mathbf{X}) = h(\bar{\mathbf{z}}) + \frac{m-1}{2} \log 2 + \frac{m}{2} \mathbb{E}_{\mathbf{z}}[\log \|\mathbf{z}\|^2]. \quad (25)$$

*Proof:* We first calculate the Jacobian determinant of the mapping  $\phi: \mathbf{z} \mapsto \mathbf{z}\mathbf{z}^T$ , which is given by  $J_\phi(\mathbf{z}) = \sqrt{\det(\mathbf{D}_\phi^T(\mathbf{z})\mathbf{D}_\phi(\mathbf{z}))}$ . By (18) and some simple algebraic manipulations, one obtains  $J_\phi(\mathbf{z}) = \sqrt{\det(2\|\mathbf{z}\|^2\mathbf{I}_m + 2\mathbf{z}\mathbf{z}^T)}$ , and further

$$\begin{aligned} J_\phi(\mathbf{z}) &= \sqrt{2^m \|\mathbf{z}\|^{2m} \det\left(\mathbf{I}_m + \frac{1}{\|\mathbf{z}\|^2} \mathbf{z}\mathbf{z}^T\right)} \\ &\stackrel{(a)}{=} \sqrt{2^m \|\mathbf{z}\|^{2m} \left(1 + \frac{\mathbf{z}^T \mathbf{z}}{\|\mathbf{z}\|^2}\right)} \\ &= \sqrt{2^{m+1} \|\mathbf{z}\|^{2m}} \\ &= 2^{\frac{m+1}{2}} \|\mathbf{z}\|^m \end{aligned} \quad (26)$$

where (a) holds due to [21, Example 1.3.24]. Because the mapping  $\phi: \mathbf{z} \mapsto \mathbf{z}\mathbf{z}^T$  is not one-to-one, we cannot directly

use Corollary 24. However, along the lines of Remark 22, we obtain

$$\begin{aligned} \mathfrak{h}^m(\mathbf{X}) &= - \int_{\mathbb{R}^m} f_z(\mathbf{z}) \log \left( \sum_{\mathbf{z}' \in \phi^{-1}(\{\phi(\mathbf{z})\})} \frac{f_z(\mathbf{z}')}{J_\phi(\mathbf{z}')} \right) d\mathcal{L}^m(\mathbf{z}). \end{aligned} \quad (27)$$

Because the  $\mathbf{z}' \in \phi^{-1}(\{\phi(\mathbf{z})\})$  are given by  $\pm \mathbf{z}$ , and because  $f_z(\mathbf{z}) + f_z(-\mathbf{z}) = 2f_{\bar{\mathbf{z}}}(\mathbf{z})$  and  $J_\phi(\mathbf{z}) = J_\phi(-\mathbf{z})$  (see (26)), eq. (27) implies

$$\begin{aligned} \mathfrak{h}^m(\mathbf{X}) &= - \int_{\mathbb{R}^m} f_z(\mathbf{z}) \log \left( 2 \frac{f_{\bar{\mathbf{z}}}(\mathbf{z})}{J_\phi(\mathbf{z})} \right) d\mathcal{L}^m(\mathbf{z}) \\ &= - \int_{\mathbb{R}^m} f_z(\mathbf{z}) (\log 2 + \log f_{\bar{\mathbf{z}}}(\mathbf{z}) - \log J_\phi(\mathbf{z})) d\mathcal{L}^m(\mathbf{z}) \\ &= -\log 2 - \int_{\mathbb{R}^m} f_z(\mathbf{z}) \log f_{\bar{\mathbf{z}}}(\mathbf{z}) d\mathcal{L}^m(\mathbf{z}) + \mathbb{E}_{\mathbf{z}}[\log J_\phi(\mathbf{z})] \\ &\stackrel{(a)}{=} -\log 2 - \frac{1}{2} \int_{\mathbb{R}^m} f_z(\mathbf{z}) \log f_{\bar{\mathbf{z}}}(\mathbf{z}) d\mathcal{L}^m(\mathbf{z}) \\ &\quad - \frac{1}{2} \int_{\mathbb{R}^m} f_z(-\mathbf{z}) \log f_{\bar{\mathbf{z}}}(\mathbf{z}) d\mathcal{L}^m(\mathbf{z}) + \mathbb{E}_{\mathbf{z}}[\log J_\phi(\mathbf{z})] \\ &= -\log 2 - \int_{\mathbb{R}^m} f_{\bar{\mathbf{z}}}(\mathbf{z}) \log f_{\bar{\mathbf{z}}}(\mathbf{z}) d\mathcal{L}^m(\mathbf{z}) + \mathbb{E}_{\mathbf{z}}[\log J_\phi(\mathbf{z})] \\ &= -\log 2 + h(\bar{\mathbf{z}}) + \mathbb{E}_{\mathbf{z}}[\log J_\phi(\mathbf{z})] \end{aligned} \quad (28)$$

where (a) holds because  $f_{\bar{\mathbf{z}}}(-\mathbf{z}) = f_{\bar{\mathbf{z}}}(\mathbf{z})$ . Inserting (26) into (28) gives (25). ■

A practically interesting special case of symmetric random matrices is constituted by the class of Wishart matrices [22]. A rank- $n$  Wishart matrix is given by  $\mathbf{W}_{n,\Sigma} \triangleq \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^T \in \mathbb{R}^{m \times m}$ , where the  $\mathbf{z}_i$ ,  $i \in \{1, \dots, n\}$  are independent and identically distributed (i.i.d.) Gaussian random variables on  $\mathbb{R}^m$  with mean  $\mathbf{0}$  and some nonsingular covariance matrix  $\Sigma$ . The differential entropy of a *full-rank* Wishart matrix (i.e.,  $n \geq m$ ), considered as a random variable in the  $m(m+1)/2$ -dimensional space of symmetric matrices, is given by [23, eq. (B.82)]

$$\begin{aligned} h(\mathbf{W}_{n,\Sigma}) &= \log \left( 2^{mn/2} \Gamma_m \left( \frac{n}{2} \right) (\det \Sigma)^{n/2} \right) \\ &\quad + \frac{mn}{2} + \frac{m-n+1}{2} \mathbb{E}_{\mathbf{z}}[\log \det(\mathbf{W}_{n,\Sigma})] \end{aligned} \quad (29)$$

where  $\Gamma_m(\cdot)$  denotes the multivariate gamma function. In our setting, full-rank Wishart matrices can be interpreted as  $m(m+1)/2$ -rectifiable random variables in the  $m^2$ -dimensional space of all  $m \times m$  matrices by considering the embedding of symmetric matrices into the space of all matrices and using Theorem 14. Using this interpretation, we can use Corollary 24 and obtain  $h(\mathbf{W}_{n,\Sigma}) = \mathfrak{h}^{m(m+1)/2}(\mathbf{W}_{n,\Sigma})$ .

The case of rank-deficient Wishart matrices, i.e.,  $n \in \{1, \dots, m-1\}$ , has not been analyzed information-theoretically so far. For simplicity, we will consider the case of rank-one Wishart matrices, i.e.,  $\mathbf{W}_{1,\Sigma} = \mathbf{z}\mathbf{z}^T \in \mathbb{R}^{m \times m}$ . The  $m$ -dimensional entropy of  $\mathbf{W}_{1,\Sigma}$  is given by (25) in Theorem 25.

Because  $\mathbf{z}$  is Gaussian with mean  $\mathbf{0}$ , we have  $\bar{\mathbf{z}} = \mathbf{z}$  in Theorem 25, so that (25) simplifies to

$$\mathfrak{h}^m(\mathbf{W}_{1,\Sigma}) = h(\mathbf{z}) + \frac{m-1}{2} \log 2 + \frac{m}{2} \mathbb{E}_{\mathbf{z}}[\log \|\mathbf{z}\|^2].$$

Again using the Gaussianity of  $\mathbf{z}$ , we obtain further

$$\begin{aligned} \mathfrak{h}^m(\mathbf{W}_{1,\Sigma}) &= \log((2\pi e)^{m/2} (\det \Sigma)^{1/2}) \\ &\quad + \frac{m-1}{2} \log 2 + \frac{m}{2} \mathbb{E}_{\mathbf{z}}[\log \|\mathbf{z}\|^2] \\ &= \log(2^{m-1/2} \pi^{m/2} (\det \Sigma)^{1/2}) \\ &\quad + \frac{m}{2} + \frac{m}{2} \mathbb{E}_{\mathbf{z}}[\log \|\mathbf{z}\|^2]. \end{aligned} \quad (30)$$

If  $\mathbf{z}$  contains independent standard normal entries, then  $\|\mathbf{z}\|^2$  is  $\chi_m^2$  distributed and  $\mathbb{E}_{\mathbf{z}}[\log \|\mathbf{z}\|^2] = \psi(m/2) + \log 2$ , where  $\psi(\cdot)$  denotes the digamma function [23, eq. (B.81)]. It is interesting to compare (30) with the differential entropy of the full-rank Wishart matrix as given by (29). Although there is a formal similarity, we emphasize that the differential entropy in (29) cannot be trivially extended to the setting  $n < m$  because neither  $\Gamma_m(\frac{n}{2})$  nor  $\log \det(\mathbf{W}_{n,\Sigma})$  is defined in this case. We conjecture that an expression similar to (30) can be derived for other rank-deficient Wishart matrices. However, as mentioned in Section III-E, the analysis of these matrices is significantly more involved and, thus, beyond the scope of this paper.

*Remark 26:* A different approach to defining an entropy for rank-deficient Wishart matrices would be to use a coordinate system on the manifold of all positive semidefinite matrices of rank  $n$  and calculate a probability density function with respect to volume elements of this manifold. Such a density was calculated for Wishart matrices in [22], and could be used for an alternative entropy definition.

## V. JOINT ENTROPY

Joint entropy is a widely used concept although it can be covered by the general concept of higher-dimensional entropy, because a pair of random variables  $(\mathbf{x}, \mathbf{y})$  with  $\mathbf{x} \in \mathbb{R}^{M_1}$  and  $\mathbf{y} \in \mathbb{R}^{M_2}$  can also be interpreted as a single random variable on  $\mathbb{R}^{M_1+M_2}$ . Thus, our concept of entropy automatically generalizes to more than one random variable. Using this interpretation, we obtain from (19) and (20) for an  $m$ -rectifiable pair of random variables  $(\mathbf{x}, \mathbf{y})$  (i.e.,  $\mu(\mathbf{x}, \mathbf{y})^{-1} \ll \mathcal{H}^m|_{\mathcal{E}}$  for an  $m$ -rectifiable set  $\mathcal{E}$ )

$$\begin{aligned} \mathfrak{h}^m(\mathbf{x}, \mathbf{y}) &\triangleq -\mathbb{E}_{(\mathbf{x}, \mathbf{y})}[\log \theta_{(\mathbf{x}, \mathbf{y})}^m(\mathbf{x}, \mathbf{y})] \\ &= -\int_{\mathbb{R}^M} \log \theta_{(\mathbf{x}, \mathbf{y})}^m(\mathbf{x}, \mathbf{y}) d\mu(\mathbf{x}, \mathbf{y})^{-1}(\mathbf{x}, \mathbf{y}) \\ &= -\int_{\mathbb{R}^M} \theta_{(\mathbf{x}, \mathbf{y})}^m(\mathbf{x}, \mathbf{y}) \log \theta_{(\mathbf{x}, \mathbf{y})}^m(\mathbf{x}, \mathbf{y}) d\mathcal{H}^m|_{\mathcal{E}}(\mathbf{x}, \mathbf{y}) \end{aligned} \quad (32)$$

with  $M = M_1 + M_2$ . However, there are still some questions to answer:

- Assuming that  $\mathbf{x}$ ,  $\mathbf{y}$ , and  $(\mathbf{x}, \mathbf{y})$  are  $m_1$ -,  $m_2$ -, and  $m$ -rectifiable, respectively, is there a relationship between the quantities  $\mathfrak{h}^{m_1}(\mathbf{x})$ ,  $\mathfrak{h}^{m_2}(\mathbf{y})$ , and  $\mathfrak{h}^m(\mathbf{x}, \mathbf{y})$  provided they exist?

- Suppose we have an  $m_1$ -rectifiable random variable  $\mathbf{x}$  and an  $m_2$ -rectifiable random variable  $\mathbf{y}$  on the same probability space. Which additional assumptions ensure that  $(\mathbf{x}, \mathbf{y})$  is  $(m_1 + m_2)$ -rectifiable?
- Conversely, suppose we have an  $m$ -rectifiable random variable  $(\mathbf{x}, \mathbf{y})$ . Which additional assumptions ensure that  $\mathbf{x}$  and  $\mathbf{y}$  are rectifiable?

In what follows, we will provide answers to these questions under appropriate conditions on the involved random variables.

One important shortcoming of Hausdorff measures (in contrast to, e.g., the Lebesgue measure) is that the product of two Hausdorff measures is in general not again a Hausdorff measure. However, our definition of the support of a rectifiable measure in Definition 8 guarantees that the product of two Hausdorff measures restricted to the respective supports is again a Hausdorff measure.

*Lemma 27:* Let  $\mathbf{x}$  be  $m_1$ -rectifiable with support  $\mathcal{E}_1$ , and let  $\mathbf{y}$  be  $m_2$ -rectifiable with support  $\mathcal{E}_2$ . Then  $\mathcal{E}_1 \times \mathcal{E}_2$  is  $(m_1 + m_2)$ -rectifiable and

$$\mathcal{H}^{m_1+m_2}|_{\mathcal{E}_1 \times \mathcal{E}_2} = \mathcal{H}^{m_1}|_{\mathcal{E}_1} \times \mathcal{H}^{m_2}|_{\mathcal{E}_2}. \quad (33)$$

*Proof:* According to Definition 11, we have  $\mathcal{E}_1 = \bigcup_{k \in \mathbb{N}} f_k(\mathcal{A}_k)$  and  $\mathcal{E}_2 = \bigcup_{k \in \mathbb{N}} g_k(\mathcal{B}_k)$  where, for  $k \in \mathbb{N}$ ,  $\mathcal{A}_k$  and  $\mathcal{B}_k$  are bounded Borel sets and  $f_k$  and  $g_k$  are Lipschitz functions that are one-to-one on  $\mathcal{A}_k$  and  $\mathcal{B}_k$ , respectively. By [20, Th. 15.1], the sets  $f_k(\mathcal{A}_k)$  and  $g_k(\mathcal{B}_k)$  are also Borel sets and, thus, [18, Th. 3.2.23] implies  $\mathcal{H}^{m_1+m_2}|_{f_k(\mathcal{A}_k) \times g_k(\mathcal{B}_k)} = \mathcal{H}^{m_1}|_{f_k(\mathcal{A}_k)} \times \mathcal{H}^{m_2}|_{g_k(\mathcal{B}_k)}$ . The result (33) then follows by the  $\sigma$ -additivity of Hausdorff measures. ■

### A. Joint Entropy for Independent Random Variables

We start our investigation of joint entropy with independent random variables. In this case, it turns out that the  $m$ -dimensional entropy is additive.

*Theorem 28:* Let  $\mathbf{x}: \Omega \rightarrow \mathbb{R}^{M_1}$  and  $\mathbf{y}: \Omega \rightarrow \mathbb{R}^{M_2}$  be independent random variables on a probability space  $(\Omega, \mathcal{G}, \mu)$ . Furthermore, let  $\mathbf{x}$  be  $m_1$ -rectifiable with support  $\mathcal{E}_1$  and let  $\mathbf{y}$  be  $m_2$ -rectifiable with support  $\mathcal{E}_2$ . Then the following properties hold:

- 1) The random variable  $(\mathbf{x}, \mathbf{y}): \Omega \rightarrow \mathbb{R}^{M_1+M_2}$  is  $(m_1 + m_2)$ -rectifiable.
- 2) The  $(m_1 + m_2)$ -dimensional Hausdorff density of  $(\mathbf{x}, \mathbf{y})$  satisfies

$$\theta_{(\mathbf{x}, \mathbf{y})}^{m_1+m_2}(\mathbf{x}, \mathbf{y}) = \theta_{\mathbf{x}}^{m_1}(\mathbf{x}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y}) \quad (34)$$

$\mathcal{H}^{m_1+m_2}$ -almost everywhere.

- 3) The set  $\mathcal{E}_1 \times \mathcal{E}_2$  is  $(m_1 + m_2)$ -rectifiable and satisfies  $\mu(\mathbf{x}, \mathbf{y})^{-1} \ll \mathcal{H}^{m_1+m_2}|_{\mathcal{E}_1 \times \mathcal{E}_2}$ .
- 4) If  $\mathfrak{h}^{m_1}(\mathbf{x})$  and  $\mathfrak{h}^{m_2}(\mathbf{y})$  are finite, then the  $(m_1 + m_2)$ -dimensional entropy of the random variable  $(\mathbf{x}, \mathbf{y})$  is given by

$$\mathfrak{h}^{m_1+m_2}(\mathbf{x}, \mathbf{y}) = \mathfrak{h}^{m_1}(\mathbf{x}) + \mathfrak{h}^{m_2}(\mathbf{y}).$$

*Proof:* See Appendix D. ■

A corollary of Theorem 28 is a result for finite sequences of independent random variables. Such sequences will be important for our discussion of typical sets in Section VIII.

*Corollary 29:* Let  $\mathbf{x}_{1:n} \triangleq (\mathbf{x}_1, \dots, \mathbf{x}_n)$  be a finite sequence of independent random variables, where  $\mathbf{x}_i \in \mathbb{R}^{M_i}$ ,  $i \in \{1, \dots, n\}$  is  $m_i$ -rectifiable with support  $\mathcal{E}_i$  and  $m_i$ -dimensional Hausdorff density  $\theta_{\mathbf{x}_i}^{m_i}$ . Then  $\mathbf{x}_{1:n}$  is an  $m$ -rectifiable random variable on  $\mathbb{R}^M$ , where  $m = \sum_{i=1}^n m_i$  and  $M = \sum_{i=1}^n M_i$ , and the set  $\mathcal{E} \triangleq \mathcal{E}_1 \times \dots \times \mathcal{E}_n$  is  $m$ -rectifiable and satisfies  $\mu(\mathbf{x}_{1:n})^{-1} \ll \mathcal{H}^m|_{\mathcal{E}}$ . Moreover, the  $m$ -dimensional Hausdorff density of  $\mathbf{x}_{1:n}$  is given by

$$\theta_{\mathbf{x}_{1:n}}^m(\mathbf{x}_{1:n}) = \prod_{i=1}^n \theta_{\mathbf{x}_i}^{m_i}(\mathbf{x}_i).$$

Finally, if  $\mathfrak{h}^{m_i}(\mathbf{x}_i)$  is finite for  $i \in \{1, \dots, n\}$ , then

$$\mathfrak{h}^m(\mathbf{x}_{1:n}) = \sum_{i=1}^n \mathfrak{h}^{m_i}(\mathbf{x}_i). \quad (35)$$

*Proof:* The corollary follows by inductively applying Theorem 28 to the two random variables  $(\mathbf{x}_1, \dots, \mathbf{x}_{i-1})$  and  $\mathbf{x}_i$ . ■

### B. Dependent Random Variables

The case of dependent random variables is more involved. The rectifiability of  $\mathbf{x}$  and  $\mathbf{y}$  does not necessarily imply the rectifiability of  $(\mathbf{x}, \mathbf{y})$  (which is expected, since the marginal distributions carry only a small part of the information carried by the joint distribution). In general, even for continuous random variables  $\mathbf{x}$  and  $\mathbf{y}$ , we cannot calculate the joint differential entropy  $h(\mathbf{x}, \mathbf{y})$  from the mere knowledge of the differential entropies  $h(\mathbf{x})$  and  $h(\mathbf{y})$ . However, it is always possible to bound the differential entropy according to [13, eq. (8.63)]

$$h(\mathbf{x}, \mathbf{y}) \leq h(\mathbf{x}) + h(\mathbf{y}). \quad (36)$$

In general, no bound resembling (36) holds for our entropy definition. The following simple setting provides a counterexample.

*Example 30:* We continue our example of a random variable on the unit circle (see Section IV-D) for the special case of a uniform distribution of  $z$  on  $[0, 2\pi)$ . From (24), we obtain

$$\mathfrak{h}^1(\mathbf{x}) = h(z) = \log(2\pi). \quad (37)$$

We can now analyze the components<sup>11</sup>  $x$  and  $y$  of the random variable  $\mathbf{x} = (x, y)^T = (\cos z, \sin z)^T$ . One can easily see that  $x$  is a continuous random variable and its probability density function is given by  $f_x(x) = 1/(\pi\sqrt{1-x^2})$ . By symmetry, the same holds for  $y$ , i.e.,  $f_y(y) = 1/(\pi\sqrt{1-y^2})$ . Basic calculus then yields for the differential entropy of  $x$  and  $y$

$$h(x) = h(y) = \log\left(\frac{\pi}{2}\right). \quad (38)$$

Since  $x$  and  $y$  are continuous random variables, it follows from Theorem 23 that  $\mathfrak{h}^1(x) = h(x)$  and  $\mathfrak{h}^1(y) = h(y)$ . Thus,

$$\mathfrak{h}^1(x) + \mathfrak{h}^1(y) = 2 \log\left(\frac{\pi}{2}\right) < \log(2\pi).$$

Comparing with (37), we see that  $\mathfrak{h}^1(\mathbf{x}, y) > \mathfrak{h}^1(x) + \mathfrak{h}^1(y)$ . ■

<sup>11</sup>To conform with the notation  $(\mathbf{x}, \mathbf{y})$  used in our treatment of joint entropy, we change the component notation from  $(x_1, x_2)^T$  to  $(x, y)^T$ .

The reason for this seemingly unintuitive behavior of our entropy are the geometric properties of the projection  $\mathbf{p}_y: \mathbb{R}^{M_1+M_2} \rightarrow \mathbb{R}^{M_2}$ ,  $\mathbf{p}_y(\mathbf{x}, \mathbf{y}) = \mathbf{y}$ , i.e., the projection of  $\mathbb{R}^{M_1+M_2}$  to the last  $M_2$  components. Although  $\mathbf{p}_y$  is linear and has a Jacobian determinant  $J_{\mathbf{p}_y}$  of 1 everywhere on  $\mathbb{R}^{M_1+M_2}$ , things get more involved once we consider  $\mathbf{p}_y$  as a mapping between rectifiable sets and want to calculate the Jacobian determinant  $J_{\mathbf{p}_y}^\mathcal{E}$  of the tangential differential of  $\mathbf{p}_y$  which maps an  $m$ -rectifiable set  $\mathcal{E} \subseteq \mathbb{R}^{M_1+M_2}$  to an  $m_2$ -rectifiable set  $\mathcal{E}_2 \subseteq \mathbb{R}^{M_2}$  [18, Sec. 3.2.16]. In this setting,  $J_{\mathbf{p}_y}^\mathcal{E}$  is not necessarily constant and may also become zero. Thus, the marginalization of an  $m$ -dimensional Hausdorff density is not as easy as the marginalization of a probability density function. The following theorem shows how to marginalize Hausdorff densities and describes the implications for  $m$ -dimensional entropy.

*Theorem 31:* Let  $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{M_1+M_2}$  be an  $m$ -rectifiable random variable ( $m \leq M_1 + M_2$ ) with  $m$ -dimensional Hausdorff density  $\theta_{(\mathbf{x}, \mathbf{y})}^m$  and support  $\mathcal{E}$ . Furthermore, let  $\tilde{\mathcal{E}}_2 \triangleq \mathbf{p}_y(\mathcal{E}) \subseteq \mathbb{R}^{M_2}$  be  $m_2$ -rectifiable ( $m_2 \leq m$ ,  $m_2 \leq M_2$ ),  $\mathcal{H}^{m_2}(\tilde{\mathcal{E}}_2) < \infty$ , and  $J_{\mathbf{p}_y}^\mathcal{E} > 0$   $\mathcal{H}^m|_{\mathcal{E}}$ -almost everywhere. Then the following properties hold:

- 1) The random variable  $\mathbf{y}$  is  $m_2$ -rectifiable.
- 2) There exists a support  $\mathcal{E}_2 \subseteq \tilde{\mathcal{E}}_2$  of  $\mathbf{y}$ .
- 3) The  $m_2$ -dimensional Hausdorff density of  $\mathbf{y}$  is given by

$$\theta_{\mathbf{y}}^{m_2}(\mathbf{y}) = \int_{\mathcal{E}(\mathbf{y})} \frac{\theta_{(\mathbf{x}, \mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{J_{\mathbf{p}_y}^\mathcal{E}(\mathbf{x}, \mathbf{y})} d\mathcal{H}^{m-m_2}(\mathbf{x}) \quad (39)$$

$\mathcal{H}^{m_2}$ -almost everywhere, where  $\mathcal{E}(\mathbf{y}) \triangleq \{\mathbf{x} \in \mathbb{R}^{M_1} : (\mathbf{x}, \mathbf{y}) \in \mathcal{E}\}$ .

- 4) An expression of the  $m_2$ -dimensional entropy of  $\mathbf{y}$  is given by

$$\mathfrak{h}^{m_2}(\mathbf{y}) = - \int_{\mathcal{E}} \theta_{(\mathbf{x}, \mathbf{y})}^m(\mathbf{x}, \mathbf{y}) \log \theta_{\mathbf{y}}^{m_2}(\mathbf{y}) d\mathcal{H}^m(\mathbf{x}, \mathbf{y}) \quad (40)$$

provided the integral on the right-hand side exists and is finite.

Under the assumptions that  $\tilde{\mathcal{E}}_1 \triangleq \mathbf{p}_x(\mathcal{E})$  is  $m_1$ -rectifiable ( $m_1 \leq m$ ,  $m_1 \leq M_1$ ),  $\mathcal{H}^{m_1}(\tilde{\mathcal{E}}_1) < \infty$ , and  $J_{\mathbf{p}_x}^\mathcal{E} > 0$   $\mathcal{H}^m|_{\mathcal{E}}$ -almost everywhere, analogous results hold for  $\mathbf{x}$ .

*Proof:* See Appendix E. ■

We will illustrate the main findings of Theorem 31 in the setting of Example 30.

*Example 32:* As in Example 30, we consider  $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^2$  uniformly distributed on the unit circle  $\mathcal{S}_1$ . By (23),  $\theta_{(\mathbf{x}, \mathbf{y})}^1(x, y) = 1/(2\pi)$   $\mathcal{H}^1$ -almost everywhere on  $\mathcal{S}_1$ . In Example 30, we already obtained  $\mathfrak{h}^1(y) = \log(\pi/2)$  (there, we used the fact that  $y$  is a continuous random variable and that, by Theorem 23,  $\mathfrak{h}^1(y) = h(y)$ ). Let us now calculate  $\mathfrak{h}^1(y)$  using Theorem 31. Note first that  $\mathbf{p}_y(\mathcal{S}_1) = [-1, 1]$ , which is 1-rectifiable and satisfies  $\mathcal{H}^1([-1, 1]) = 2 < \infty$ . Next, we calculate the Jacobian determinant  $J_{\mathbf{p}_y}^{\mathcal{S}_1}(x, y)$ . Consider an arbitrary point on the unit circle, which can always be expressed as  $(\pm\sqrt{1-y^2}, \pm y)$  with  $y \in [0, 1]$ . At that point, the projection  $\mathbf{p}_y$  restricted to the tangent space of  $\mathcal{S}_1$  can be shown to amount to a multiplication by the factor  $\sqrt{1-y^2}$ .

Thus,  $J_{p_y}^{S_1}(\pm\sqrt{1-y^2}, \pm y) = \sqrt{1-y^2}$ . Hence, we obtain from (40)

$$\begin{aligned}
h^1(y) &= - \int_{S_1} \theta_{(x,y)}^1(x, y) \\
&\quad \times \log \left( \int_{S_1^{(y)}} \frac{\theta_{(x,y)}^1(\tilde{x}, y)}{J_{p_y}^{S_1}(\tilde{x}, y)} d\mathcal{H}^{1-1}(\tilde{x}) \right) d\mathcal{H}^1(x, y) \\
&= - \int_{S_1} \frac{1}{2\pi} \log \left( \int_{S_1^{(y)}} \frac{\frac{1}{2\pi}}{\sqrt{1-y^2}} d\mathcal{H}^0(\tilde{x}) \right) d\mathcal{H}^1(x, y) \\
&\stackrel{(a)}{=} - \frac{1}{2\pi} \int_{S_1} \log \left( \sum_{\tilde{x} \in S_1^{(y)}} \frac{\frac{1}{2\pi}}{\sqrt{1-y^2}} \right) d\mathcal{H}^1(x, y) \\
&\stackrel{(b)}{=} - \frac{1}{2\pi} \int_{S_1} \log \left( 2 \frac{\frac{1}{2\pi}}{\sqrt{1-y^2}} \right) d\mathcal{H}^1(x, y) \\
&= - \frac{1}{2\pi} \int_0^{2\pi} \log \left( \frac{1}{\pi |\cos(\phi)|} \right) d\phi \\
&= \log \left( \frac{\pi}{2} \right)
\end{aligned} \tag{41}$$

where (a) holds because  $\mathcal{H}^0$  is the counting measure and (b) holds because  $S_1^{(y)} = \{x \in \mathbb{R} : (x, y) \in S_1\} = \{\sqrt{1-y^2}, -\sqrt{1-y^2}\}$  contains two points for all  $y \in (-1, 1)$ . Note that our above result for  $h^1(y)$  coincides with the result previously obtained in Example 30. ■

### C. Product-Compatible Random Variables

There are special settings in which  $m$ -dimensional entropy more closely matches the behavior we know from (differential) entropy. In these cases, the three random variables  $\mathbf{x}$ ,  $\mathbf{y}$ , and  $(\mathbf{x}, \mathbf{y})$  are rectifiable with “matching” dimensions, and we will see that an inequality similar to (36) holds.

**Definition 33:** Let  $\mathbf{x}$  be an  $m_1$ -rectifiable random variable on  $\mathbb{R}^{M_1}$  with support  $\mathcal{E}_1$ , and let  $\mathbf{y}$  be an  $m_2$ -rectifiable random variable on  $\mathbb{R}^{M_2}$  with support  $\mathcal{E}_2$ . The random variables  $\mathbf{x}$  and  $\mathbf{y}$  are called *product-compatible* if  $(\mathbf{x}, \mathbf{y})$  is an  $(m_1 + m_2)$ -rectifiable random variable on  $\mathbb{R}^{M_1+M_2}$ .

It is easy to see that for product-compatible random variables  $\mathbf{x}$  and  $\mathbf{y}$ ,  $\mu(\mathbf{x}, \mathbf{y})^{-1} \ll \mathcal{H}^{m_1+m_2}|_{\mathcal{E}_1 \times \mathcal{E}_2}$ . Thus, by Property 4 in Corollary 12, there exists a support  $\mathcal{E} \subseteq \mathcal{E}_1 \times \mathcal{E}_2$ .

The most important part of Definition 33 is that the dimensions of  $\mathbf{x}$  and  $\mathbf{y}$  add up to the joint dimension of  $(\mathbf{x}, \mathbf{y})$ . Note that this was not the case in Example 32, where  $\mathbf{x}$  and  $\mathbf{y}$  “shared” the dimension  $m = 1$  of  $(\mathbf{x}, \mathbf{y})$ . A simple example of product-compatible random variables is the case of an  $m_1$ -rectifiable random variable  $\mathbf{x}$  and an independent  $m_2$ -rectifiable random variable  $\mathbf{y}$ . Indeed, by Theorem 28,  $(\mathbf{x}, \mathbf{y})$  is  $(m_1 + m_2)$ -rectifiable.

Another example of product-compatible random variables can be deduced from Theorem 31. Let  $(\mathbf{x}, \mathbf{y})$  be  $(m_1 + m_2)$ -rectifiable. Assume that  $\tilde{\mathcal{E}}_2 \triangleq p_y(\mathcal{E}) \subseteq \mathbb{R}^{M_2}$  is  $m_2$ -rectifiable,  $\mathcal{H}^{m_2}(\tilde{\mathcal{E}}_2) < \infty$ , and  $J_{p_y}^{\mathcal{E}} > 0$   $\mathcal{H}^m|_{\mathcal{E}}$ -almost everywhere. Furthermore, assume that  $\mathcal{E}_1 \triangleq p_x(\mathcal{E}) \subseteq \mathbb{R}^{M_1}$  is  $m_1$ -rectifiable,  $\mathcal{H}^{m_1}(\mathcal{E}_1) < \infty$ , and  $J_{p_x}^{\mathcal{E}} > 0$   $\mathcal{H}^m|_{\mathcal{E}}$ -almost everywhere. By

Theorem 31,  $\mathbf{x}$  is  $m_1$ -rectifiable and  $\mathbf{y}$  is  $m_2$ -rectifiable. Thus,  $\mathbf{x}$  and  $\mathbf{y}$  are product-compatible.

The setting of product-compatible random variables will be especially important for our discussion of mutual information in Section VII. However, already for joint entropy, we obtain some useful results.

**Theorem 34:** Let  $\mathbf{x}$  be an  $m_1$ -rectifiable random variable on  $\mathbb{R}^{M_1}$  with support  $\mathcal{E}_1$ , and let  $\mathbf{y}$  be an  $m_2$ -rectifiable random variable on  $\mathbb{R}^{M_2}$  with support  $\mathcal{E}_2$ . Furthermore, let  $\mathbf{x}$  and  $\mathbf{y}$  be product-compatible. Denote by  $\theta_{(\mathbf{x}, \mathbf{y})}^{m_1+m_2}$  the  $(m_1 + m_2)$ -dimensional Hausdorff density of  $(\mathbf{x}, \mathbf{y})$  and by  $\mathcal{E} \subseteq \mathcal{E}_1 \times \mathcal{E}_2$  a support of  $(\mathbf{x}, \mathbf{y})$ . Then the following properties hold:

- 1) The  $m_2$ -dimensional Hausdorff density of  $\mathbf{y}$  is given by

$$\theta_{\mathbf{y}}^{m_2}(\mathbf{y}) = \int_{\mathcal{E}_1} \theta_{(\mathbf{x}, \mathbf{y})}^{m_1+m_2}(\mathbf{x}, \mathbf{y}) d\mathcal{H}^{m_1}(\mathbf{x})$$

$\mathcal{H}^{m_2}$ -almost everywhere.

- 2) An expression of the  $m_2$ -dimensional entropy of  $\mathbf{y}$  is given by

$$\begin{aligned}
h^{m_2}(\mathbf{y}) &= - \int_{\mathcal{E}} \theta_{(\mathbf{x}, \mathbf{y})}^{m_1+m_2}(\mathbf{x}, \mathbf{y}) \\
&\quad \times \log \theta_{\mathbf{y}}^{m_2}(\mathbf{y}) d\mathcal{H}^{m_1+m_2}(\mathbf{x}, \mathbf{y})
\end{aligned}$$

provided the integral on the right-hand side exists and is finite.

Due to symmetry, analogous properties hold for  $\theta_{\mathbf{x}}^{m_1}$  and  $h^{m_1}(\mathbf{x})$ .

*Proof:* The proof follows along the lines of the proof of Theorem 31 in Appendix E. However, due to the product-compatibility of  $\mathbf{x}$  and  $\mathbf{y}$ , one can use Fubini’s theorem in place of (110). ■

For product-compatible random variables, also the inequality  $h^{m_1+m_2}(\mathbf{x}, \mathbf{y}) \leq h^{m_1}(\mathbf{x}) + h^{m_2}(\mathbf{y})$  holds. However, the proof of this inequality will be much easier once we considered the mutual information between rectifiable random variables. Thus, we postpone a formal presentation of the inequality to Corollary 47 in Section VII.

## VI. CONDITIONAL ENTROPY

In contrast to joint entropy, conditional entropy is a nontrivial extension of entropy. We would like to define the entropy for a random variable  $\mathbf{x}$  on  $\mathbb{R}^{M_1}$  under the condition that a dependent random variable  $\mathbf{y}$  on  $\mathbb{R}^{M_2}$  is known. For discrete and—under appropriate assumptions—for continuous random variables, the distribution of  $(\mathbf{x}|\mathbf{y} = \mathbf{y})$  is well defined and so is the associated entropy  $H(\mathbf{x}|\mathbf{y} = \mathbf{y})$  or differential entropy  $h(\mathbf{x}|\mathbf{y} = \mathbf{y})$ . Averaging over all  $\mathbf{y}$  then results in the well-known definitions of conditional entropy  $H(\mathbf{x}|\mathbf{y})$ , involving only the probability mass functions  $p_{(\mathbf{x}, \mathbf{y})}$  and  $p_{\mathbf{y}}$ , or of conditional differential entropy  $h(\mathbf{x}|\mathbf{y})$ , involving only the probability density functions  $f_{(\mathbf{x}, \mathbf{y})}$  and  $f_{\mathbf{y}}$ . Indeed, if  $\mathbf{x}$  and

$\mathbf{y}$  are discrete random variables, we have

$$\begin{aligned} H(\mathbf{x}|\mathbf{y}) &= \sum_{j \in \mathbb{N}} p_{\mathbf{y}}(\mathbf{y}_j) H(\mathbf{x}|\mathbf{y} = \mathbf{y}_j) \\ &= - \sum_{i,j \in \mathbb{N}} p_{(\mathbf{x},\mathbf{y})}(x_i, \mathbf{y}_j) \log \left( \frac{p_{(\mathbf{x},\mathbf{y})}(x_i, \mathbf{y}_j)}{p_{\mathbf{y}}(\mathbf{y}_j)} \right) \\ &= -\mathbb{E}_{(\mathbf{x},\mathbf{y})} \left[ \log \left( \frac{p_{(\mathbf{x},\mathbf{y})}(\mathbf{x}, \mathbf{y})}{p_{\mathbf{y}}(\mathbf{y})} \right) \right] \end{aligned} \quad (42)$$

and, if  $\mathbf{x}$  and  $\mathbf{y}$  are continuous random variables, we have

$$\begin{aligned} h(\mathbf{x}|\mathbf{y}) &= \int_{\mathbb{R}^{M_2}} f_{\mathbf{y}}(\mathbf{y}) h(\mathbf{x}|\mathbf{y} = \mathbf{y}) d\mathbf{y} \\ &= - \int_{\mathbb{R}^{M_1+M_2}} f_{(\mathbf{x},\mathbf{y})}(\mathbf{x}, \mathbf{y}) \log \left( \frac{f_{(\mathbf{x},\mathbf{y})}(\mathbf{x}, \mathbf{y})}{f_{\mathbf{y}}(\mathbf{y})} \right) d(\mathbf{x}, \mathbf{y}) \\ &= -\mathbb{E}_{(\mathbf{x},\mathbf{y})} \left[ \log \left( \frac{f_{(\mathbf{x},\mathbf{y})}(\mathbf{x}, \mathbf{y})}{f_{\mathbf{y}}(\mathbf{y})} \right) \right]. \end{aligned} \quad (43)$$

A straightforward generalization to rectifiable measures would be to mimic the right-hand sides of (42) and (43) using Hausdorff densities. However, it will turn out that this naive approach is only partly correct: due to the geometric subtleties of the projection discussed in Section V-B, we may have to include a correction term that reflects the geometry of the conditioning process.

#### A. Conditional Probability

For general random variables  $\mathbf{x}$  and  $\mathbf{y}$ , we recall the concept of conditional probabilities, which can be summarized as follows (a detailed account can be found in [24, Ch. 5]): For a pair of random variables  $(\mathbf{x}, \mathbf{y})$  on  $\mathbb{R}^{M_1+M_2}$ , there exists a regular conditional probability  $\Pr\{\mathbf{x} \in \mathcal{A} | \mathbf{y} = \mathbf{y}\}$ , i.e., for each measurable set  $\mathcal{A} \subseteq \mathbb{R}^{M_1}$ , the function  $\mathbf{y} \mapsto \Pr\{\mathbf{x} \in \mathcal{A} | \mathbf{y} = \mathbf{y}\}$  is measurable and  $\Pr\{\mathbf{x} \in \cdot | \mathbf{y} = \mathbf{y}\}$  defines a probability measure for each  $\mathbf{y} \in \mathbb{R}^{M_2}$ . Furthermore, the regular conditional probability  $\Pr\{\mathbf{x} \in \mathcal{A} | \mathbf{y} = \mathbf{y}\}$  satisfies

$$\Pr\{(\mathbf{x}, \mathbf{y}) \in \mathcal{A}_1 \times \mathcal{A}_2\} = \int_{\mathcal{A}_2} \Pr\{\mathbf{x} \in \mathcal{A}_1 | \mathbf{y} = \mathbf{y}\} d\mu_{\mathbf{y}}^{-1}(\mathbf{y}). \quad (44)$$

The regular conditional probability  $\Pr\{\mathbf{x} \in \mathcal{A} | \mathbf{y} = \mathbf{y}\}$  involved in (44) is not unique. Nevertheless, we can still use (44) in a definition of conditional entropy because any version of the regular conditional probability satisfies (44). For the remainder of this section, we consider a fixed version of the regular conditional probability  $\Pr\{\mathbf{x} \in \mathcal{A} | \mathbf{y} = \mathbf{y}\}$ .

#### B. Definition of Conditional Entropy

In order to define a conditional entropy  $\mathfrak{h}^{m-m_2}(\mathbf{x}|\mathbf{y})$ , we first show that  $\Pr\{\mathbf{x} \in \cdot | \mathbf{y} = \mathbf{y}\}$  is a rectifiable measure. The next theorem establishes sufficient conditions such that  $\Pr\{\mathbf{x} \in \cdot | \mathbf{y} = \mathbf{y}\}$  is rectifiable for almost every  $\mathbf{y}$ . As before, we denote by  $\mathbf{p}_{\mathbf{y}}: \mathbb{R}^{M_1+M_2} \rightarrow \mathbb{R}^{M_2}$  the projection of  $\mathbb{R}^{M_1+M_2}$  to the last  $M_2$  components, i.e.,  $\mathbf{p}_{\mathbf{y}}(\mathbf{x}, \mathbf{y}) = \mathbf{y}$ .

**Theorem 35:** Let  $(\mathbf{x}, \mathbf{y})$  be an  $m$ -rectifiable random variable on  $\mathbb{R}^{M_1+M_2}$  with  $m$ -dimensional Hausdorff density  $\theta_{(\mathbf{x},\mathbf{y})}^m$  and support  $\mathcal{E}$ . Furthermore, let  $\tilde{\mathcal{E}}_2 \triangleq \mathbf{p}_{\mathbf{y}}(\mathcal{E}) \subseteq \mathbb{R}^{M_2}$  be  $m_2$ -rectifiable ( $m_2 \leq m$ ,  $m_2 \leq M_2$ ,  $m - m_2 \leq M_1$ ),

$\mathcal{H}^{m_2}(\tilde{\mathcal{E}}_2) < \infty$ , and  $J_{\mathbf{p}_{\mathbf{y}}}^{\mathcal{E}} > 0$   $\mathcal{H}^m|_{\mathcal{E}}$ -almost everywhere. Then the following properties hold:

- 1) The measure  $\Pr\{\mathbf{x} \in \cdot | \mathbf{y} = \mathbf{y}\}$  is  $(m - m_2)$ -rectifiable for  $\mathcal{H}^{m_2}|_{\mathcal{E}_2}$ -almost every  $\mathbf{y} \in \mathbb{R}^{M_2}$ , where  $\mathcal{E}_2 \subseteq \tilde{\mathcal{E}}_2$  is a support<sup>12</sup> of  $\mathbf{y}$ .
- 2) The  $(m - m_2)$ -dimensional Hausdorff density of the measure  $\Pr\{\mathbf{x} \in \cdot | \mathbf{y} = \mathbf{y}\}$  is given by

$$\theta_{\Pr\{\mathbf{x} \in \cdot | \mathbf{y} = \mathbf{y}\}}^{m-m_2}(\mathbf{x}) = \frac{\theta_{(\mathbf{x},\mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{J_{\mathbf{p}_{\mathbf{y}}}^{\mathcal{E}}(\mathbf{x}, \mathbf{y}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y})} \quad (45)$$

$\mathcal{H}^{m-m_2}|_{\mathcal{E}(\mathbf{y})}$ -almost everywhere, for  $\mathcal{H}^{m_2}|_{\mathcal{E}_2}$ -almost every  $\mathbf{y} \in \mathbb{R}^{M_2}$ . Here, as before,  $\mathcal{E}(\mathbf{y}) \triangleq \{\mathbf{x} \in \mathbb{R}^{M_1} : (\mathbf{x}, \mathbf{y}) \in \mathcal{E}\}$ .

*Proof:* See Appendix F. ■

As for joint entropy, the case of product-compatible random variables (see Definition 33) is of special interest and results in a more intuitive characterization of the Hausdorff density of  $\Pr\{\mathbf{x} \in \cdot | \mathbf{y} = \mathbf{y}\}$ .

**Theorem 36:** Let  $\mathbf{x}$  be an  $m_1$ -rectifiable random variable on  $\mathbb{R}^{M_1}$  with support  $\mathcal{E}_1$ , and let  $\mathbf{y}$  be an  $m_2$ -rectifiable random variable on  $\mathbb{R}^{M_2}$  with support  $\mathcal{E}_2$ . Furthermore, let  $\mathbf{x}$  and  $\mathbf{y}$  be product-compatible. Then the following properties hold:

- 1) The measure  $\Pr\{\mathbf{x} \in \cdot | \mathbf{y} = \mathbf{y}\}$  is  $m_1$ -rectifiable for  $\mathcal{H}^{m_2}|_{\mathcal{E}_2}$ -almost every  $\mathbf{y} \in \mathbb{R}^{M_2}$ .
- 2) The  $m_1$ -dimensional Hausdorff density of  $\Pr\{\mathbf{x} \in \cdot | \mathbf{y} = \mathbf{y}\}$  is given by

$$\theta_{\Pr\{\mathbf{x} \in \cdot | \mathbf{y} = \mathbf{y}\}}^{m_1}(\mathbf{x}) = \frac{\theta_{(\mathbf{x},\mathbf{y})}^{m_1+m_2}(\mathbf{x}, \mathbf{y})}{\theta_{\mathbf{y}}^{m_2}(\mathbf{y})} \quad (46)$$

$\mathcal{H}^{m_1}|_{\mathcal{E}_1}$ -almost everywhere, for  $\mathcal{H}^{m_2}|_{\mathcal{E}_2}$ -almost every  $\mathbf{y} \in \mathbb{R}^{M_2}$ .

*Proof:* The proof follows along the lines of the proof of Theorem 35 in Appendix F. However, due to the product-compatibility of  $\mathbf{x}$  and  $\mathbf{y}$ , one can use Fubini's theorem in place of (110). ■

Note that Theorems 35 and 36 hold for any version of the regular conditional probability  $\Pr\{\mathbf{x} \in \mathcal{A} | \mathbf{y} = \mathbf{y}\}$ . However, for different versions, the statement “for  $\mathcal{H}^{m_2}|_{\mathcal{E}_2}$ -almost every  $\mathbf{y} \in \mathbb{R}^{M_2}$ ” may refer to different sets of  $\mathcal{H}^{m_2}|_{\mathcal{E}_2}$ -measure zero; e.g., (45) may hold for different  $\mathbf{y} \in \mathbb{R}^{M_2}$ . Thus, results that are independent of the version of the regular conditional probability can only be obtained if we can avoid these “almost everywhere”-statements. To this end, we will define conditional entropy as an expectation over  $\mathbf{y}$ .

**Definition 37:** Let  $(\mathbf{x}, \mathbf{y})$  be an  $m$ -rectifiable random variable on  $\mathbb{R}^{M_1+M_2}$  such that  $\mathbf{y}$  is  $m_2$ -rectifiable with  $m_2$ -dimensional Hausdorff density  $\theta_{\mathbf{y}}^{m_2}$  and support  $\mathcal{E}_2$ . The *conditional entropy of  $\mathbf{x}$  given  $\mathbf{y}$*  is defined as<sup>13</sup>

$$\begin{aligned} \mathfrak{h}^{m-m_2}(\mathbf{x}|\mathbf{y}) &\triangleq - \int_{\mathcal{E}_2} \theta_{\mathbf{y}}^{m_2}(\mathbf{y}) \int_{\mathcal{E}(\mathbf{y})} \theta_{\Pr\{\mathbf{x} \in \cdot | \mathbf{y} = \mathbf{y}\}}^{m-m_2}(\mathbf{x}) \\ &\quad \times \log \theta_{\Pr\{\mathbf{x} \in \cdot | \mathbf{y} = \mathbf{y}\}}^{m-m_2}(\mathbf{x}) d\mathcal{H}^{m-m_2}(\mathbf{x}) d\mathcal{H}^{m_2}(\mathbf{y}) \end{aligned} \quad (47)$$

<sup>12</sup>By Theorem 31, the random variable  $\mathbf{y}$  is  $m_2$ -rectifiable with Hausdorff density  $\theta_{\mathbf{y}}^{m_2}$  (given by (39)) and some support  $\mathcal{E}_2 \subseteq \tilde{\mathcal{E}}_2$ .

<sup>13</sup>The inner integral in (47) can be intuitively interpreted as an entropy  $\mathfrak{h}^{m-m_2}(\mathbf{x}|\mathbf{y} = \mathbf{y})$ . However, such an entropy is not well defined in general and depends on the choice of the conditional probability.

provided the right-hand side of (47) exists and coincides for all versions of the regular conditional probability  $\Pr\{\mathbf{x} \in \mathcal{A} \mid \mathbf{y} = \mathbf{y}\}$ .

*Remark 38:* For independent random variables  $\mathbf{x}$  and  $\mathbf{y}$ , inserting (34) into (46) implies that  $\theta_{\Pr\{\mathbf{x} \in \cdot \mid \mathbf{y} = \mathbf{y}\}}^{m_1}(\mathbf{x}) = \theta_{\mathbf{x}}^{m_1}(\mathbf{x})$ . Thus, (47) reduces to  $\mathfrak{h}^{m_1}(\mathbf{x} \mid \mathbf{y}) = \mathfrak{h}^{m_1}(\mathbf{x})$ .

The following theorem gives a characterization of conditional entropy and sufficient conditions for (47) to be well-defined in the sense that the right-hand side of (47) coincides for all versions of the regular conditional probability  $\Pr\{\mathbf{x} \in \mathcal{A} \mid \mathbf{y} = \mathbf{y}\}$ .

*Theorem 39:* Let  $(\mathbf{x}, \mathbf{y})$  be an  $m$ -rectifiable random variable on  $\mathbb{R}^{M_1+M_2}$  with  $m$ -dimensional Hausdorff density  $\theta_{(\mathbf{x}, \mathbf{y})}^m$  and support  $\mathcal{E}$ . Furthermore, let  $\mathcal{E}_2 \triangleq \mathbf{p}_{\mathbf{y}}(\mathcal{E})$  be  $m_2$ -rectifiable,  $\mathcal{H}^{m_2}(\mathcal{E}_2) < \infty$ , and  $J_{\mathbf{p}_{\mathbf{y}}}^{\mathcal{E}} > 0$   $\mathcal{H}^m|_{\mathcal{E}}$ -almost everywhere. Then

$$\mathfrak{h}^{m-m_2}(\mathbf{x} \mid \mathbf{y}) = -\mathbb{E}_{(\mathbf{x}, \mathbf{y})} \left[ \log \left( \frac{\theta_{(\mathbf{x}, \mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{\theta_{\mathbf{y}}^{m_2}(\mathbf{y})} \right) \right] + \mathbb{E}_{(\mathbf{x}, \mathbf{y})} [\log J_{\mathbf{p}_{\mathbf{y}}}^{\mathcal{E}}(\mathbf{x}, \mathbf{y})] \quad (48)$$

provided the right-hand side of (48) exists and is finite.

*Proof:* See Appendix G. ■

Note the difference between (48) and the expressions (42) and (43) of  $H(\mathbf{x} \mid \mathbf{y})$  and  $h(\mathbf{x} \mid \mathbf{y})$ , respectively: in the case of rectifiable random variables, we generally have to include the geometric correction term  $\mathbb{E}_{(\mathbf{x}, \mathbf{y})} [\log J_{\mathbf{p}_{\mathbf{y}}}^{\mathcal{E}}(\mathbf{x}, \mathbf{y})]$ . However, we will show next that, in the special case of product-compatible rectifiable random variables, this correction term does not appear.

*Theorem 40:* Let the  $m_1$ -rectifiable random variable  $\mathbf{x}$  on  $\mathbb{R}^{M_1}$  and the  $m_2$ -rectifiable random variable  $\mathbf{y}$  on  $\mathbb{R}^{M_2}$  be product-compatible. Then

$$\mathfrak{h}^{m_1}(\mathbf{x} \mid \mathbf{y}) = -\mathbb{E}_{(\mathbf{x}, \mathbf{y})} \left[ \log \left( \frac{\theta_{(\mathbf{x}, \mathbf{y})}^{m_1+m_2}(\mathbf{x}, \mathbf{y})}{\theta_{\mathbf{y}}^{m_2}(\mathbf{y})} \right) \right] \quad (49)$$

provided the right-hand side of (49) exists and is finite.

*Proof:* The proof follows along the lines of the proof of Theorem 39 in Appendix G. However, due to the product-compatibility of  $\mathbf{x}$  and  $\mathbf{y}$ , one can use Fubini's theorem in place of (110). ■

### C. Chain Rule for Rectifiable Random Variables

As in the case of entropy and differential entropy, we can give a chain rule for  $m$ -dimensional entropy.

*Theorem 41:* Let  $(\mathbf{x}, \mathbf{y})$  be an  $m$ -rectifiable random variable on  $\mathbb{R}^{M_1+M_2}$  with  $m$ -dimensional Hausdorff density  $\theta_{(\mathbf{x}, \mathbf{y})}^m$  and support  $\mathcal{E}$ . Furthermore, let  $\mathcal{E}_2 \triangleq \mathbf{p}_{\mathbf{y}}(\mathcal{E})$  be  $m_2$ -rectifiable,  $\mathcal{H}^{m_2}(\mathcal{E}_2) < \infty$ , and  $J_{\mathbf{p}_{\mathbf{y}}}^{\mathcal{E}} > 0$   $\mathcal{H}^m|_{\mathcal{E}}$ -almost everywhere. Then

$$\mathfrak{h}^m(\mathbf{x}, \mathbf{y}) = \mathfrak{h}^{m_2}(\mathbf{y}) + \mathfrak{h}^{m-m_2}(\mathbf{x} \mid \mathbf{y}) - \mathbb{E}_{(\mathbf{x}, \mathbf{y})} [\log J_{\mathbf{p}_{\mathbf{y}}}^{\mathcal{E}}(\mathbf{x}, \mathbf{y})] \quad (50)$$

provided the corresponding integrals exist and are finite.

*Proof:* By the definition of  $\mathfrak{h}^m(\mathbf{x}, \mathbf{y})$  in (31) and the definition of  $\mathfrak{h}^{m_2}(\mathbf{y})$  in (19), we have

$$\begin{aligned} \mathfrak{h}^m(\mathbf{x}, \mathbf{y}) &= \mathfrak{h}^{m_2}(\mathbf{y}) + \mathbb{E}_{(\mathbf{x}, \mathbf{y})} [\log J_{\mathbf{p}_{\mathbf{y}}}^{\mathcal{E}}(\mathbf{x}, \mathbf{y})] \\ &= -\mathbb{E}_{(\mathbf{x}, \mathbf{y})} [\log \theta_{(\mathbf{x}, \mathbf{y})}^m(\mathbf{x}, \mathbf{y})] + \mathbb{E}_{\mathbf{y}} [\log \theta_{\mathbf{y}}^{m_2}(\mathbf{y})] \\ &\quad + \mathbb{E}_{(\mathbf{x}, \mathbf{y})} [\log J_{\mathbf{p}_{\mathbf{y}}}^{\mathcal{E}}(\mathbf{x}, \mathbf{y})] \\ &= -\mathbb{E}_{(\mathbf{x}, \mathbf{y})} \left[ \log \left( \frac{\theta_{(\mathbf{x}, \mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{\theta_{\mathbf{y}}^{m_2}(\mathbf{y})} \right) \right] + \mathbb{E}_{(\mathbf{x}, \mathbf{y})} [\log J_{\mathbf{p}_{\mathbf{y}}}^{\mathcal{E}}(\mathbf{x}, \mathbf{y})]. \end{aligned} \quad (51)$$

Because we assumed in the theorem that the integrals corresponding to the terms on the left-hand side of (51) are finite, the right-hand side of (51) is also finite. By (48), the right-hand side of (51) equals  $\mathfrak{h}^{m-m_2}(\mathbf{x} \mid \mathbf{y})$ . Thus, (50) holds. ■

Next, we continue Examples 30 and 32 from Section V-B. We will see that the geometric correction term in the chain rule,  $\mathbb{E}_{(\mathbf{x}, \mathbf{y})} [\log J_{\mathbf{p}_{\mathbf{y}}}^{\mathcal{E}}(\mathbf{x}, \mathbf{y})]$ , is indeed necessary.

*Example 42:* As in Examples 30 and 32, we consider  $(x, y) \in \mathbb{R}^2$  uniformly distributed on the unit circle  $\mathcal{S}_1$ , i.e.,  $\theta_{(x, y)}^1(x, y) = 1/(2\pi)$   $\mathcal{H}^1$ -almost everywhere on  $\mathcal{S}_1$ . According to (41),

$$\mathfrak{h}^1(y) = \log \left( \frac{\pi}{2} \right) \quad (52)$$

and according to (37),

$$\mathfrak{h}^1(x, y) = \log(2\pi). \quad (53)$$

To calculate the conditional entropy  $\mathfrak{h}^0(x \mid y)$  (note that  $m - m_2 = 1 - 1 = 0$ ), we consider the regular conditional probability  $\Pr\{x \in \mathcal{A} \mid y = y\}$ . It is easy to see that one possible version of  $\Pr\{x \in \mathcal{A} \mid y = y\}$  is the following: for  $y \in (-1, 1)$ ,  $\Pr\{x = x \mid y = y\} = 1/2$  for  $x = \pm\sqrt{1-y^2}$  and  $\Pr\{x \in \mathcal{A} \mid y = y\} = 0$  if  $\pm\sqrt{1-y^2} \notin \mathcal{A}$ . The probabilities for  $|y| \geq 1$  are irrelevant because  $\Pr\{y \notin (-1, 1)\} = 0$ . Hence, by (47), we obtain

$$\begin{aligned} \mathfrak{h}^0(x \mid y) &= - \int_{(-1, 1)} \theta_y^1(y) \int_{\{\pm\sqrt{1-y^2}\}} \frac{1}{2} \log \frac{1}{2} d\mathcal{H}^0(x) d\mathcal{H}^1(y) \\ &= - \int_{(-1, 1)} \theta_y^1(y) \log \frac{1}{2} d\mathcal{H}^1(y) \\ &= \log 2. \end{aligned} \quad (54)$$

This differs from  $\mathfrak{h}^1(x, y) - \mathfrak{h}^1(y) = \log(2\pi) - \log(\pi/2)$ , and therefore the conjecture that there holds a chain rule without a correction term is wrong. To calculate the correction term, which according to (50) is given by  $\mathbb{E}_{(\mathbf{x}, \mathbf{y})} [\log J_{\mathbf{p}_{\mathbf{y}}}^{\mathcal{S}_1}(\mathbf{x}, \mathbf{y})]$ , we recall from Example 32 that  $J_{\mathbf{p}_{\mathbf{y}}}^{\mathcal{S}_1}(\pm\sqrt{1-y^2}, \pm y) = \sqrt{1-y^2}$  or, more conveniently,  $J_{\mathbf{p}_{\mathbf{y}}}^{\mathcal{S}_1}(\cos \phi, \sin \phi) = |\cos \phi|$ . Thus, we obtain

$$\begin{aligned} \mathbb{E}_{(\mathbf{x}, \mathbf{y})} [\log J_{\mathbf{p}_{\mathbf{y}}}^{\mathcal{S}_1}(\mathbf{x}, \mathbf{y})] &= \int_{\mathcal{S}_1} \frac{1}{2\pi} \log J_{\mathbf{p}_{\mathbf{y}}}^{\mathcal{S}_1}(x, y) d\mathcal{H}^1(x, y) \\ &= \int_0^{2\pi} \frac{1}{2\pi} \log |\cos \phi| d\phi \\ &= -\log 2. \end{aligned} \quad (55)$$

We finally verify that (55) is consistent with the chain rule (50). Starting from (53), we obtain

$$\begin{aligned} \mathfrak{h}^1(\mathbf{x}, \mathbf{y}) &= \log(2\pi) \\ &= \log\left(\frac{\pi}{2}\right) + \log 2 - (-\log 2) \\ &= \mathfrak{h}^1(\mathbf{y}) + \mathfrak{h}^0(\mathbf{x} | \mathbf{y}) - \mathbb{E}_{(\mathbf{x}, \mathbf{y})} [\log J_{\mathbf{p}_y}^{S_1}(\mathbf{x}, \mathbf{y})] \end{aligned}$$

where the final expansion is obtained by using (52), (54), and (55). ■

Example 42 also provides a counterexample to the rule “conditioning does not increase entropy,” which holds for the entropy of discrete random variables and the differential entropy of continuous random variables. Indeed, comparing (38) and (54), we see that for the components of a uniform distribution on the unit circle, we have  $\mathfrak{h}^1(\mathbf{x}) < \mathfrak{h}^0(\mathbf{x} | \mathbf{y})$ . However, as we will see in Corollary 47 in Section VII, this is only due to a “reduction of dimensions”: if  $\mathbf{x}$  and  $\mathbf{y}$  are product-compatible, which implies that  $\mathfrak{h}^{m_1}(\mathbf{x})$  and  $\mathfrak{h}^{m-m_2}(\mathbf{x} | \mathbf{y})$  are of the same dimension  $m_1 = m - m_2$ , conditioning will indeed not increase entropy, i.e.,  $\mathfrak{h}^{m_1}(\mathbf{x} | \mathbf{y}) \leq \mathfrak{h}^{m_1}(\mathbf{x})$ . Also the chain rule (50) reduces to its traditional form, as stated next.

*Theorem 43:* Let the  $m_1$ -rectifiable random variable  $\mathbf{x}$  on  $\mathbb{R}^{M_1}$  and the  $m_2$ -rectifiable random variable  $\mathbf{y}$  on  $\mathbb{R}^{M_2}$  be product-compatible. Then

$$\mathfrak{h}^{m_1+m_2}(\mathbf{x}, \mathbf{y}) = \mathfrak{h}^{m_2}(\mathbf{y}) + \mathfrak{h}^{m_1}(\mathbf{x} | \mathbf{y}) \quad (56)$$

provided the entropies  $\mathfrak{h}^{m_1+m_2}(\mathbf{x}, \mathbf{y})$  and  $\mathfrak{h}^{m_2}(\mathbf{y})$  exist and are finite.

*Proof:* By the definition of  $\mathfrak{h}^{m_1+m_2}(\mathbf{x}, \mathbf{y})$  in (31) and the definition of  $\mathfrak{h}^{m_2}(\mathbf{y})$  in (19), we have

$$\begin{aligned} \mathfrak{h}^{m_1+m_2}(\mathbf{x}, \mathbf{y}) - \mathfrak{h}^{m_2}(\mathbf{y}) &= -\mathbb{E}_{(\mathbf{x}, \mathbf{y})} [\log \theta_{(\mathbf{x}, \mathbf{y})}^{m_1+m_2}(\mathbf{x}, \mathbf{y})] + \mathbb{E}_{\mathbf{y}} [\log \theta_{\mathbf{y}}^{m_2}(\mathbf{y})] \\ &= -\mathbb{E}_{(\mathbf{x}, \mathbf{y})} \left[ \log \left( \frac{\theta_{(\mathbf{x}, \mathbf{y})}^{m_1+m_2}(\mathbf{x}, \mathbf{y})}{\theta_{\mathbf{y}}^{m_2}(\mathbf{y})} \right) \right]. \end{aligned} \quad (57)$$

By (49), the right-hand side of (57) equals  $\mathfrak{h}^{m_1}(\mathbf{x} | \mathbf{y})$ . Thus, (56) holds. ■

Using an induction argument, we can extend the chain rule (56) to a sequence of random variables.

*Corollary 44:* Let  $\mathbf{x}_{1:n} \triangleq (\mathbf{x}_1, \dots, \mathbf{x}_n)$  be a sequence of random variables where each  $\mathbf{x}_i \in \mathbb{R}^{M_i}$  is  $m_i$ -rectifiable. Assume that  $\mathbf{x}_{1:i-1}$  and  $\mathbf{x}_i$  are product-compatible for  $i \in \{2, \dots, n\}$ . Then

$$\mathfrak{h}^m(\mathbf{x}_{1:n}) = \mathfrak{h}^{m_1}(\mathbf{x}_1) + \sum_{i=2}^n \mathfrak{h}^{m_i}(\mathbf{x}_i | \mathbf{x}_{1:i-1}) \quad (58)$$

with  $m = \sum_{i=1}^n m_i$ , provided the corresponding integrals exist and are finite.

We note that, consistently with Remark 38, (35) is a special case of (58).

## VII. MUTUAL INFORMATION

The basic definition of mutual information is for discrete random variables  $\mathbf{x}$  and  $\mathbf{y}$  with probability mass functions  $p_{\mathbf{x}}(\mathbf{x}_i)$  and  $p_{\mathbf{y}}(\mathbf{y}_j)$  and joint probability mass function  $p_{(\mathbf{x}, \mathbf{y})}(\mathbf{x}_i, \mathbf{y}_j)$ . The mutual information between  $\mathbf{x}$  and  $\mathbf{y}$  is given by [13, eq. (2.28)]

$$I(\mathbf{x}; \mathbf{y}) \triangleq \sum_{i,j} p_{(\mathbf{x}, \mathbf{y})}(\mathbf{x}_i, \mathbf{y}_j) \log \left( \frac{p_{(\mathbf{x}, \mathbf{y})}(\mathbf{x}_i, \mathbf{y}_j)}{p_{\mathbf{x}}(\mathbf{x}_i)p_{\mathbf{y}}(\mathbf{y}_j)} \right). \quad (59)$$

However, mutual information is also defined between arbitrary random variables  $\mathbf{x}$  and  $\mathbf{y}$  on a common probability space. This definition is based on (59) and quantizations  $[\mathbf{x}]_{\Omega}$  and  $[\mathbf{y}]_{\mathfrak{R}}$  [13, eq. (8.54)]. We recall from Section II-A that for a measurable, finite partition  $\Omega = \{\mathcal{A}_1, \dots, \mathcal{A}_N\}$  of  $\mathbb{R}^{M_1}$  (i.e.,  $\mathbb{R}^{M_1} = \bigcup_{i=1}^N \mathcal{A}_i$  with  $\mathcal{A}_i \in \Omega$  mutually disjoint and measurable), the quantization  $[\mathbf{x}]_{\Omega} \in \{1, \dots, N\}$  is defined as the discrete random variable with probability mass function  $p_{[\mathbf{x}]_{\Omega}}(i) = \Pr\{[\mathbf{x}]_{\Omega} = i\} = \Pr\{\mathbf{x} \in \mathcal{A}_i\}$  for  $i \in \{1, \dots, N\}$ .

*Definition 45* ([13, eq. (8.54)]): Let  $\mathbf{x}: \Omega \rightarrow \mathbb{R}^{M_1}$  and  $\mathbf{y}: \Omega \rightarrow \mathbb{R}^{M_2}$  be random variables on a common probability space  $(\Omega, \mathfrak{S}, \mu)$ . The mutual information between  $\mathbf{x}$  and  $\mathbf{y}$  is defined as

$$I(\mathbf{x}; \mathbf{y}) \triangleq \sup_{\Omega, \mathfrak{R}} I([\mathbf{x}]_{\Omega}; [\mathbf{y}]_{\mathfrak{R}})$$

where the supremum is taken over all measurable, finite partitions  $\Omega$  of  $\mathbb{R}^{M_1}$  and  $\mathfrak{R}$  of  $\mathbb{R}^{M_2}$ .

The Gelfand-Yaglom-Perez theorem [25, Lem. 5.2.3] provides an expression of mutual information in terms of Radon-Nikodym derivatives: for random variables  $\mathbf{x}: \Omega \rightarrow \mathbb{R}^{M_1}$  and  $\mathbf{y}: \Omega \rightarrow \mathbb{R}^{M_2}$  on a common probability space  $(\Omega, \mathfrak{S}, \mu)$ ,

$$I(\mathbf{x}; \mathbf{y}) = \int_{\mathbb{R}^{M_1+M_2}} \log \left( \frac{d\mu(\mathbf{x}, \mathbf{y})^{-1}}{d(\mu_{\mathbf{x}}^{-1} \times \mu_{\mathbf{y}}^{-1})}(x, y) \right) \times d\mu(\mathbf{x}, \mathbf{y})^{-1}(x, y) \quad (60)$$

if  $\mu(\mathbf{x}, \mathbf{y})^{-1} \ll \mu_{\mathbf{x}}^{-1} \times \mu_{\mathbf{y}}^{-1}$ , and

$$I(\mathbf{x}; \mathbf{y}) = \infty \quad (61)$$

if  $\mu(\mathbf{x}, \mathbf{y})^{-1} \not\ll \mu_{\mathbf{x}}^{-1} \times \mu_{\mathbf{y}}^{-1}$ .

For the special cases of discrete and continuous random variables, there exist expressions of mutual information in terms of entropy and differential entropy, respectively. We will extend these expressions to the case of rectifiable random variables. The resulting generalization will involve the entropies  $\mathfrak{h}^{m_1}(\mathbf{x})$ ,  $\mathfrak{h}^{m_2}(\mathbf{y})$ , and  $\mathfrak{h}^m(\mathbf{x}, \mathbf{y})$ .

*Theorem 46:* Let  $\mathbf{x}$  be an  $m_1$ -rectifiable random variable with support  $\mathcal{E}_1 \subseteq \mathbb{R}^{M_1}$ , let  $\mathbf{y}$  be an  $m_2$ -rectifiable random variable with support  $\mathcal{E}_2 \subseteq \mathbb{R}^{M_2}$ , and let  $(\mathbf{x}, \mathbf{y})$  be  $m$ -rectifiable with support  $\mathcal{E} \subseteq \mathcal{E}_1 \times \mathcal{E}_2$ . The mutual information  $I(\mathbf{x}; \mathbf{y})$  satisfies:

- 1) If  $\mathbf{x}$  and  $\mathbf{y}$  are product-compatible (i.e.,  $m = m_1 + m_2$ ), then

$$\begin{aligned} I(\mathbf{x}; \mathbf{y}) &= \int_{\mathcal{E}} \theta_{(\mathbf{x}, \mathbf{y})}^m(x, y) \\ &\quad \times \log \left( \frac{\theta_{(\mathbf{x}, \mathbf{y})}^m(x, y)}{\theta_{\mathbf{x}}^{m_1}(x)\theta_{\mathbf{y}}^{m_2}(y)} \right) d\mathcal{H}^m(x, y). \end{aligned} \quad (62)$$

Furthermore,

$$I(\mathbf{x}; \mathbf{y}) = \mathfrak{h}^{m_1}(\mathbf{x}) + \mathfrak{h}^{m_2}(\mathbf{y}) - \mathfrak{h}^m(\mathbf{x}, \mathbf{y}) \quad (63)$$

and

$$I(\mathbf{x}; \mathbf{y}) = \mathfrak{h}^{m_1}(\mathbf{x}) - \mathfrak{h}^{m_1}(\mathbf{x} | \mathbf{y}) = \mathfrak{h}^{m_2}(\mathbf{y}) - \mathfrak{h}^{m_2}(\mathbf{y} | \mathbf{x}) \quad (64)$$

provided the entropies  $\mathfrak{h}^{m_1}(\mathbf{x})$ ,  $\mathfrak{h}^{m_2}(\mathbf{y})$ , and  $\mathfrak{h}^m(\mathbf{x}, \mathbf{y})$  exist and are finite.

2) If  $m < m_1 + m_2$ , then  $I(\mathbf{x}; \mathbf{y}) = \infty$ .

*Proof:* See Appendix H. ■

In Theorem 46, the case  $m < m_1 + m_2$  can be interpreted as  $\mathbf{x}$  and  $\mathbf{y}$  “sharing” at least one dimension. In a communication scenario, this would imply that it is possible to reconstruct an at least one-dimensional component of  $\mathbf{x}$  from  $\mathbf{y}$  (and, also, to reconstruct an at least one-dimensional component of  $\mathbf{y}$  from  $\mathbf{x}$ ). Thus, an infinite amount of information could be transmitted over a channel  $\mathbf{x} \rightarrow \mathbf{y}$  (or  $\mathbf{y} \rightarrow \mathbf{x}$ ). This is consistent with our result that  $I(\mathbf{x}; \mathbf{y}) = \infty$ .

A corollary of Theorem 46 states that for product-compatible random variables, we can upper-bound the joint entropy by the sum of the individual entropies and prove that conditioning does not increase entropy.

*Corollary 47:* Let the  $m_1$ -rectifiable random variable  $\mathbf{x}$  on  $\mathbb{R}^{M_1}$  and the  $m_2$ -rectifiable random variable  $\mathbf{y}$  on  $\mathbb{R}^{M_2}$  be product-compatible. Then

$$\mathfrak{h}^{m_1+m_2}(\mathbf{x}, \mathbf{y}) \leq \mathfrak{h}^{m_1}(\mathbf{x}) + \mathfrak{h}^{m_2}(\mathbf{y}) \quad (65)$$

and

$$\mathfrak{h}^{m_1}(\mathbf{x} | \mathbf{y}) \leq \mathfrak{h}^{m_1}(\mathbf{x}) \quad (66)$$

provided the entropies  $\mathfrak{h}^{m_1}(\mathbf{x})$ ,  $\mathfrak{h}^{m_2}(\mathbf{y})$ , and  $\mathfrak{h}^{m_1+m_2}(\mathbf{x}, \mathbf{y})$  exist and are finite.

*Proof:* The inequality (65) follows from (63) and the nonnegativity of mutual information. Similarly, (66) follows from (64) and the nonnegativity of mutual information. ■

### VIII. ASYMPTOTIC EQUIPARTITION PROPERTY

Similar to classical entropy and differential entropy, the  $m$ -dimensional entropy  $\mathfrak{h}^m(\mathbf{x})$  satisfies an asymptotic equipartition property (AEP). Let us consider a sequence  $\mathbf{x}_{1:n} \triangleq (\mathbf{x}_1, \dots, \mathbf{x}_n)$  of i.i.d. random variables  $\mathbf{x}_i$ . Our main findings are similar to the discrete and continuous cases: based on  $\mathfrak{h}^m(\mathbf{x})$ , we define sets  $\mathcal{A}_\varepsilon^{(n)}$  of typical sequences  $\mathbf{x}_{1:n}$  and show that, for sufficiently large  $n$ , a random sequence  $\mathbf{x}_{1:n}$  belongs to  $\mathcal{A}_\varepsilon^{(n)}$  with probability arbitrarily close to one. Furthermore, we obtain upper and lower bounds on the size of  $\mathcal{A}_\varepsilon^{(n)}$  given by  $e^{n(\mathfrak{h}^m(\mathbf{x})+\varepsilon)}$  and  $(1-\delta)e^{n(\mathfrak{h}^m(\mathbf{x})-\varepsilon)}$ , respectively. In the case of classical entropy and differential entropy, these properties are useful in the proof of various coding theorems because they allow us to consider only typical sequences.

Our analysis follows the steps in [13, Sec. 8.2]. However, whereas in the discrete case the size of a set of sequences  $\mathbf{x}_{1:n}$  is measured by its cardinality and in the continuous case by its Lebesgue measure, in the present case of  $m$ -rectifiable random variables  $\mathbf{x}_i$ , we resort to the Hausdorff measure.

*Lemma 48:* Let  $\mathbf{x}_{1:n} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$  be a sequence of i.i.d.  $m$ -rectifiable random variables  $\mathbf{x}_i$  on  $\mathbb{R}^M$ , where each  $\mathbf{x}_i$

has  $m$ -dimensional Hausdorff density  $\theta_{\mathbf{x}}^m$  and  $m$ -dimensional entropy  $\mathfrak{h}^m(\mathbf{x})$ . The random variable  $-(1/n) \sum_{i=1}^n \log \theta_{\mathbf{x}}^m(\mathbf{x}_i)$  converges to  $\mathfrak{h}^m(\mathbf{x})$  in probability, i.e., for any  $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \Pr \left\{ \left| -\frac{1}{n} \sum_{i=1}^n \log \theta_{\mathbf{x}}^m(\mathbf{x}_i) - \mathfrak{h}^m(\mathbf{x}) \right| > \varepsilon \right\} = 0.$$

*Proof:* By (19), we have  $\mathfrak{h}^m(\mathbf{x}) = -\mathbb{E}_{\mathbf{x}}[\log \theta_{\mathbf{x}}^m(\mathbf{x})]$ , and by the weak law of large numbers, the sample mean  $-(1/n) \sum_{i=1}^n \log \theta_{\mathbf{x}}^m(\mathbf{x}_i)$  converges in probability to the expectation  $-\mathbb{E}_{\mathbf{x}}[\log \theta_{\mathbf{x}}^m(\mathbf{x})]$ . ■

We can define typical sets in the usual way [13, Sec. 8.2].

*Definition 49:* Let  $\mathbf{x}$  be an  $m$ -rectifiable random variable on  $\mathbb{R}^M$  with support  $\mathcal{E}$  and  $m$ -dimensional Hausdorff density  $\theta_{\mathbf{x}}^m$ . For  $\varepsilon > 0$  and  $n \in \mathbb{N}$ , the  $\varepsilon$ -typical set  $\mathcal{A}_\varepsilon^{(n)} \subseteq \mathbb{R}^{nM}$  is defined as

$$\mathcal{A}_\varepsilon^{(n)} \triangleq \left\{ \mathbf{x}_{1:n} \in \mathcal{E}^n : \left| -\frac{1}{n} \sum_{i=1}^n \log \theta_{\mathbf{x}}^m(\mathbf{x}_i) - \mathfrak{h}^m(\mathbf{x}) \right| \leq \varepsilon \right\}.$$

The AEP for sequences of  $m$ -rectifiable random variables is expressed by the following central result.

*Theorem 50:* Let  $\mathbf{x}_{1:n} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$  be a sequence of i.i.d.  $m$ -rectifiable random variables  $\mathbf{x}_i$  on  $\mathbb{R}^M$ , where each  $\mathbf{x}_i$  has  $m$ -dimensional Hausdorff density  $\theta_{\mathbf{x}}^m$ , support  $\mathcal{E}$ , and  $m$ -dimensional entropy  $\mathfrak{h}^m(\mathbf{x})$ . Then the typical set  $\mathcal{A}_\varepsilon^{(n)}$  satisfies the following properties.

1) For  $\delta > 0$  and  $n$  sufficiently large,

$$\Pr\{\mathbf{x}_{1:n} \in \mathcal{A}_\varepsilon^{(n)}\} > 1 - \delta.$$

2) For all  $n \in \mathbb{N}$ ,

$$\mathcal{H}^{nm}(\mathcal{A}_\varepsilon^{(n)}) \leq e^{n(\mathfrak{h}^m(\mathbf{x})+\varepsilon)}.$$

3) For  $\delta > 0$  and  $n$  sufficiently large,

$$\mathcal{H}^{nm}(\mathcal{A}_\varepsilon^{(n)}) > (1 - \delta)e^{n(\mathfrak{h}^m(\mathbf{x})-\varepsilon)}.$$

*Proof:* The proof is similar to that in the continuous case [13, Th. 8.2.2], however with the Lebesgue measure replaced by the Hausdorff measure. ■

### IX. ENTROPY BOUNDS ON EXPECTED CODEWORD LENGTH

A well-known result for discrete random variables is a connection between the minimal expected codeword length of an instantaneous source code and the entropy of the random variable [13, Th. 5.4.1]. More specifically, let  $\mathbf{x}$  be a discrete random variable on  $\mathbb{R}^M$  with possible realizations  $\{\mathbf{x}_i : i \in \mathcal{I}\}$ . In variable-length source coding, a one-to-one function  $f: \{\mathbf{x}_i : i \in \mathcal{I}\} \rightarrow \{0, 1\}^*$ , where  $\{0, 1\}^*$  denotes the set of all finite-length binary sequences, is used to represent each realization  $\mathbf{x}_i$  by a finite-length binary sequence  $s_i = f(\mathbf{x}_i)$ . This code is instantaneous (or prefix free) if no  $f(\mathbf{x}_i)$  coincides with the first bits of another  $f(\mathbf{x}_j)$ . The expected binary codeword length is defined as

$$L_f(\mathbf{x}) \triangleq \mathbb{E}_{\mathbf{x}}[\ell(f(\mathbf{x}))]$$

where  $\ell(s)$  denotes the length of a binary sequence  $s \in \{0, 1\}^*$ . The minimal expected binary codeword length  $L^*(\mathbf{x})$



is defined as the minimum of  $L_f(\mathbf{x})$  over the set of all possible instantaneous codes  $f$ . By [13, Th. 5.4.1],  $L^*(\mathbf{x})$  satisfies<sup>14</sup>

$$H(\mathbf{x}) \text{ld } e \leq L^*(\mathbf{x}) < H(\mathbf{x}) \text{ld } e + 1. \quad (67)$$

#### A. Expected Codeword Length of an Integer-Dimensional Random Variable

For a nondiscrete  $m$ -rectifiable random variable  $\mathbf{x}$  (i.e.,  $m \geq 1$ ), a one-to-one code of finite expected codeword length does not exist. However, quantizations of  $\mathbf{x}$  can be encoded using finite-length binary sequences. We will present results for the minimal expected codeword length of constrained quantizations of  $\mathbf{x}$ .

**Definition 51:** Let  $\mathcal{E} \subseteq \mathbb{R}^M$  be an  $m$ -rectifiable set. Furthermore, let  $\Omega = \{\mathcal{A}_1, \dots, \mathcal{A}_N\}$  be a finite  $\mathcal{H}^m$ -measurable partition of  $\mathcal{E}$ , i.e., all sets  $\mathcal{A}_i$  are mutually disjoint and  $\mathcal{H}^m$ -measurable, and  $\bigcup_{i=1}^N \mathcal{A}_i = \mathcal{E}$ . The partition  $\Omega$  is said to be an  $(m, \delta)$ -partition of  $\mathcal{E}$  if  $\mathcal{H}^m(\mathcal{A}_i) \leq \delta$  for all  $i \in \{1, \dots, N\}$ . The set of all  $(m, \delta)$ -partitions of  $\mathcal{E}$  is denoted  $\mathfrak{P}_{m, \delta}^{(\mathcal{E})}$ .

Note that the definition of an  $(m, \delta)$ -partition of an  $m$ -rectifiable set  $\mathcal{E}$  does not involve a distortion function. On the one hand, this is convenient because we do not have to argue about a good distortion measure. On the other hand, the points in a set  $\mathcal{A}_i$  of a partition  $\Omega \in \mathfrak{P}_{m, \delta}^{(\mathcal{E})}$  are not necessarily “close” to each other; in fact,  $\mathcal{A}_i$  is not even necessarily connected. Thus, although the partitions in  $\mathfrak{P}_{m, \delta}^{(\mathcal{E})}$  consist of measure-theoretically small sets, these sets might be considered large in terms of specific distortion measures.

In what follows, we will consider the quantized random variable  $[\mathbf{x}]_\Omega$  for  $\Omega \in \mathfrak{P}_{m, \delta}^{(\mathcal{E})}$ . We recall that  $[\mathbf{x}]_\Omega$  is the discrete random variable such that  $\Pr\{[\mathbf{x}]_\Omega = i\} = \Pr\{\mathbf{x} \in \mathcal{A}_i\}$  for  $i \in \{1, \dots, N\}$ . Due to the interpretation of  $\mathfrak{h}^m(\mathbf{x})$  as a generalized entropy (cf. Remark 19), we can use [12, eq. (1.8)] to obtain the following result.

**Lemma 52:** Let  $\mathbf{x}$  be an  $m$ -rectifiable random variable, i.e.,  $\mu\mathbf{x}^{-1} \ll \mathcal{H}^m|_{\mathcal{E}}$  for an  $m$ -rectifiable set  $\mathcal{E} \subseteq \mathbb{R}^M$ , with  $m \geq 1$  and  $\mathcal{H}^m(\mathcal{E}) < \infty$ . Let  $\mathfrak{P}_{m, \infty}^{(\mathcal{E})}$  denote the set of all finite,  $\mathcal{H}^m$ -measurable partitions of  $\mathcal{E}$ . Then

$$\begin{aligned} \mathfrak{h}^m(\mathbf{x}) &= \inf_{\Omega \in \mathfrak{P}_{m, \infty}^{(\mathcal{E})}} \left( - \sum_{\mathcal{A} \in \Omega} \mu\mathbf{x}^{-1}(\mathcal{A}) \log \left( \frac{\mu\mathbf{x}^{-1}(\mathcal{A})}{\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A})} \right) \right) \quad (68) \\ &= \inf_{\Omega \in \mathfrak{P}_{m, \infty}^{(\mathcal{E})}} \left( H([\mathbf{x}]_\Omega) + \sum_{\mathcal{A} \in \Omega} \mu\mathbf{x}^{-1}(\mathcal{A}) \log \mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}) \right). \quad (69) \end{aligned}$$

*Proof:* See Appendix I. ■

The terms in (69) give an interesting interpretation of  $m$ -dimensional entropy. Looking for a quantization that minimizes the first term,  $H([\mathbf{x}]_\Omega)$ , corresponds to minimizing the amount of data required to represent this quantization. Of course, the minimum is simply obtained for the partition  $\Omega = \{\mathcal{E}\}$ , which gives  $H([\mathbf{x}]_\Omega) = 0$ . But in (69), we also have an additional term that penalizes a bad “resolution” of

the quantization: if the quantized random variable  $[\mathbf{x}]_\Omega$  is with high probability—corresponding to  $\mu\mathbf{x}^{-1}(\mathcal{A})$  being large—in a large quantization set  $\mathcal{A}$ , then this is penalized by the term  $\mu\mathbf{x}^{-1}(\mathcal{A}) \log \mathcal{H}^m|_{\mathcal{E}}(\mathcal{A})$ . Thus, (69) shows that  $m$ -dimensional entropy can be interpreted in terms of a tradeoff between fine resolution and efficient representation.

We now turn to a generalization of (67) to rectifiable random variables.

**Theorem 53:** Let  $\mathbf{x}$  be an  $m$ -rectifiable random variable, i.e.,  $\mu\mathbf{x}^{-1} \ll \mathcal{H}^m|_{\mathcal{E}}$  for an  $m$ -rectifiable set  $\mathcal{E} \subseteq \mathbb{R}^M$ , with  $m \geq 1$  and  $\mathcal{H}^m(\mathcal{E}) < \infty$ . For any  $\Omega \in \mathfrak{P}_{m, \delta}^{(\mathcal{E})}$ , the minimal expected binary codeword length of the quantized random variable  $[\mathbf{x}]_\Omega$  satisfies

$$L^*([\mathbf{x}]_\Omega) \geq \mathfrak{h}^m(\mathbf{x}) \text{ld } e - \text{ld } \delta. \quad (70)$$

Furthermore, for each  $\varepsilon > 0$ , there exists  $\delta_\varepsilon > 0$  such that the following holds: for each  $\delta \in (0, \delta_\varepsilon)$ , there exists a partition  $\Omega_\delta \in \mathfrak{P}_{m, \delta}^{(\mathcal{E})}$  such that

$$L^*([\mathbf{x}]_{\Omega_\delta}) < \mathfrak{h}^m(\mathbf{x}) \text{ld } e - \text{ld } \delta + 1 + \varepsilon. \quad (71)$$

*Proof:* See Appendix J. We note that the proof is based on (67) and the expression of  $\mathfrak{h}^m(\mathbf{x})$  given in (69). ■

The lower bound (70) shows the following: if we want a quantization  $\Omega$  of  $\mathbf{x}$  with good resolution (in the sense that  $\mathcal{H}^m(\mathcal{A}) \leq \delta$  for all  $\mathcal{A} \in \Omega$ ), then we have to use at least  $\mathfrak{h}^m(\mathbf{x}) \text{ld } e - \text{ld } \delta$  bits to represent this quantized random variable using an instantaneous code. However, by the upper bound (71), we know that for a sufficiently fine resolution (i.e.,  $\delta < \delta_\varepsilon$ ), that resolution  $\delta$  can be achieved by using at most  $1 + \varepsilon$  additional bits (in addition to the lower bound  $\mathfrak{h}^m(\mathbf{x}) \text{ld } e - \text{ld } \delta$ ).

#### B. Expected Codeword Length of Sequences of Integer-Dimensional Random Variables

We will now apply Theorem 53 to sequences of i.i.d. random variables. To this end, we consider quantizations of an entire sequence,  $[\mathbf{x}_{1:n}]_\Omega = [(\mathbf{x}_1, \dots, \mathbf{x}_n)]_\Omega$  with<sup>15</sup>  $\Omega \in \mathfrak{P}_{nm, \delta^n}^{(\mathcal{E}^n)}$ . We denote by

$$L_n^*([\mathbf{x}_{1:n}]_\Omega) \triangleq \frac{L^*([\mathbf{x}_{1:n}]_\Omega)}{n} \quad (72)$$

the minimal expected binary codeword length per source symbol.

**Corollary 54:** Let  $\mathbf{x}_{1:n} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$  be a sequence of i.i.d.  $m$ -rectifiable random variables ( $m \geq 1$ ) on  $\mathbb{R}^M$  with  $m$ -dimensional entropy  $\mathfrak{h}^m(\mathbf{x})$  and support  $\mathcal{E}$  satisfying  $\mathcal{H}^m(\mathcal{E}) < \infty$ . Then, for each  $\varepsilon > 0$ , there exists  $\delta_\varepsilon > 0$  such that the following holds: for each  $\delta \in (0, \delta_\varepsilon)$ , there exists a partition  $\Omega \in \mathfrak{P}_{nm, \delta^n}^{(\mathcal{E}^n)}$  such that the minimal expected binary codeword length per source symbol satisfies

$$\mathfrak{h}^m(\mathbf{x}) \text{ld } e - \text{ld } \delta \leq L_n^*([\mathbf{x}_{1:n}]_\Omega) \leq \mathfrak{h}^m(\mathbf{x}) \text{ld } e - \text{ld } \delta + \frac{1 + \varepsilon}{n}. \quad (73)$$

<sup>15</sup>We choose partitions  $\Omega$  of resolution  $\delta^n$ , i.e., the sets  $\mathcal{A} \in \Omega$  satisfy  $\mathcal{H}^{nm}(\mathcal{A}) \leq \delta^n$ . This choice is made for consistency with the case of partitions  $\Omega$  of  $\mathcal{E}^n$  that are constructed as products of sets  $\mathcal{A}_i$  in  $\Omega_1 \in \mathfrak{P}_{m, \delta}^{(\mathcal{E})}$ . More specifically, for  $\mathcal{A} = \mathcal{A}_1 \times \dots \times \mathcal{A}_n$  with  $\mathcal{A}_i \in \Omega_1$ , we have  $\mathcal{H}^m(\mathcal{A}_i) \leq \delta$  and  $\mathcal{H}^{nm}(\mathcal{A}) \leq \delta^n$  and the sets  $\mathcal{A}$  cover  $\mathcal{E}^n$ , i.e.,  $\Omega \triangleq \{\mathcal{A} = \mathcal{A}_1 \times \dots \times \mathcal{A}_n : \mathcal{A}_i \in \Omega_1\} \in \mathfrak{P}_{nm, \delta^n}^{(\mathcal{E}^n)}$ .

<sup>14</sup>The factor  $\text{ld } e$  appears because we defined entropy using the natural logarithm.

*Proof:* By Corollary 29, the random variable  $\mathbf{x}_{1:n}$  is  $nm$ -rectifiable with  $\mu(\mathbf{x}_{1:n})^{-1} \ll \mathcal{H}^m|_{\mathcal{E}^n}$  and  $nm$ -dimensional entropy  $\mathfrak{h}^{nm}(\mathbf{x}_{1:n}) = n\mathfrak{h}^m(\mathbf{x})$ . Thus, by Theorem 53, there exists  $\hat{\delta}_\varepsilon > 0$  such that the following holds:

(\*) For all  $\hat{\delta} \in (0, \hat{\delta}_\varepsilon)$ , there exists a partition  $\Omega \in \mathfrak{P}_{nm, \hat{\delta}}^{(\mathcal{E}^n)}$  such that

$$\begin{aligned} n\mathfrak{h}^m(\mathbf{x}) \text{ld } e - \text{ld } \hat{\delta} &\leq L^*([\mathbf{x}_{1:n}]_\Omega) \\ &< n\mathfrak{h}^m(\mathbf{x}) \text{ld } e - \text{ld } \hat{\delta} + 1 + \varepsilon. \end{aligned}$$

Define  $\delta_\varepsilon \triangleq \hat{\delta}_\varepsilon^{1/n}$  and let  $\delta \in (0, \delta_\varepsilon)$ . We have that  $\delta \in (0, \delta_\varepsilon)$  is equivalent to  $\delta^n \in (0, \hat{\delta}_\varepsilon)$ . Thus, by (\*) for the specific case  $\hat{\delta} = \delta^n$ , there exists a partition  $\Omega \in \mathfrak{P}_{nm, \delta^n}^{(\mathcal{E}^n)}$  such that

$$\begin{aligned} n\mathfrak{h}^m(\mathbf{x}) \text{ld } e - \text{ld } \delta^n &\leq L^*([\mathbf{x}_{1:n}]_\Omega) \\ &< n\mathfrak{h}^m(\mathbf{x}) \text{ld } e - \text{ld } \delta^n + 1 + \varepsilon. \end{aligned}$$

Dividing by  $n$  and using (72) gives (73). ■

Corollary 54 shows that the upper bound on the expected codeword length per source symbol becomes closer to the lower bound  $\mathfrak{h}^m(\mathbf{x}) \text{ld } e - \text{ld } \delta$  if we are allowed to quantize and code entire sequences. However, note that using the quantization  $\Omega \in \mathfrak{P}_{nm, \delta^n}^{(\mathcal{E}^n)}$  of the joint random variable  $\mathbf{x}_{1:n}$ , it is not guaranteed that we can reconstruct each  $\mathbf{x}_i$  to within a set  $\mathcal{A}_i$  satisfying  $\mathcal{H}^m(\mathcal{A}_i) \leq \delta$ . All we know is that each  $\mathcal{A} \in \Omega$  satisfies  $\mathcal{H}^{nm}(\mathcal{A}) \leq \delta^n$ , i.e., the overall resolution of the sequence is good, but the resolution of each individual source symbol is not necessarily good too.

## X. SHANNON LOWER BOUND FOR INTEGER-DIMENSIONAL SOURCES

As a second application of the proposed entropy definition, we present a lower bound on the rate-distortion (RD) function of integer-dimensional sources. The RD function for a source  $\mathbf{x}$  and a distortion function  $d(\cdot, \cdot)$  is defined as [4, eq. (4.1.3)]

$$R(D) \triangleq \inf_{\mathbb{E}_{(\mathbf{x}, \mathbf{y})}[d(\mathbf{x}, \mathbf{y})] \leq D} I(\mathbf{x}; \mathbf{y})$$

for  $D \geq 0$ , where the constrained infimum is taken over all joint probability distributions of  $(\mathbf{x}, \mathbf{y})$  with the given probability distribution of  $\mathbf{x}$  as the first marginal. We will consider throughout this section a source random variable  $\mathbf{x}$  on  $\mathbb{R}^M$  and a translation invariant distortion function  $d(\cdot, \cdot)$  on  $\mathbb{R}^M \times \mathbb{R}^M$ , i.e.,  $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{x} - \mathbf{y}, \mathbf{0})$  for all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^M$ . Furthermore, we assume that  $d(\cdot, \cdot)$  satisfies  $\inf_{\mathbf{y} \in \mathbb{R}^M} d(\mathbf{x}, \mathbf{y}) = 0$  for each  $\mathbf{x} \in \mathbb{R}^M$ . We also assume that there exist  $D \geq 0$  such that  $R(D)$  is finite, and we denote by  $D_0$  the infimum of these  $D$ . Finally, we assume that there exists a finite set  $\mathcal{B} \subseteq \mathbb{R}^M$  such that  $\mathbb{E}_{\mathbf{x}}[\min_{\mathbf{y} \in \mathcal{B}} d(\mathbf{x}, \mathbf{y})] < \infty$ . This assumption guarantees that there exists a finite quantization of  $\mathbf{x}$  with bounded expected distortion. Under these standard assumptions, we have the following characterization of the RD function [26, Th. 2.3]: For each  $D > D_0$ ,

$$R(D) = \max_{s \geq 0} \max_{\alpha_s(\cdot)} (-sD + \mathbb{E}_{\mathbf{x}}[\log \alpha_s(\mathbf{x})]) \quad (74)$$

where the second maximization is with respect to all functions<sup>16</sup>  $\alpha_s: \mathbb{R}^M \rightarrow (0, \infty)$  satisfying

$$\mathbb{E}_{\mathbf{x}}[\alpha_s(\mathbf{x})e^{-sd(\mathbf{x}, \mathbf{y})}] \leq 1 \quad (75)$$

for each  $\mathbf{y} \in \mathbb{R}^M$ .

### A. Shannon Lower Bound

The most common form of the traditional Shannon lower bound [4, Sec. 4.3] for a *discrete* source  $\mathbf{x}$  is the following inequality

$$R(D) \geq H(\mathbf{x}) - \max H(\mathbf{w}) \quad (76)$$

where the maximum is taken over all discrete random variables  $\mathbf{w}$  whose expected distortion relative to  $\mathbf{0}$  is equal to  $D$ , i.e.,  $\mathbb{E}_{\mathbf{w}}[d(\mathbf{w}, \mathbf{0})] = D$ . An important aspect of the bound (76) is that the contribution of the source  $\mathbf{x}$  and the contribution of the distortion function  $d(\cdot, \cdot)$  and distortion  $D$  become separated. For a fixed distortion function and a given distortion, we can calculate  $\max H(\mathbf{w})$  and then use the bound (76) for different sources  $\mathbf{x}$  simply by calculating their entropy  $H(\mathbf{x})$ .

For a *continuous* random variable  $\mathbf{x}$  on  $\mathbb{R}^M$ , a bound similar to (76) can also be derived under certain assumptions. However, it is more convenient to state the continuous Shannon lower bound in the following parametric form (i.e., involving a parameter  $s \geq 0$ ) [4, Sec. 4.6]

$$R(D) \geq h(\mathbf{x}) - sD - \log \tilde{\gamma}(s) \quad (77)$$

where

$$\tilde{\gamma}(s) \triangleq \int_{\mathbb{R}^M} e^{-sd(\mathbf{x}, \mathbf{0})} d\mathcal{L}^M(\mathbf{x}) \quad (78)$$

and (77) holds for all  $s \geq 0$ . The right-hand side of (77) can be maximized with respect to  $s$ , and it turns out that [4, Lem. 4.6.2]

$$\min_{s \geq 0} (sD + \log \tilde{\gamma}(s)) = \max h(\mathbf{w})$$

where the maximum is taken over all continuous random variables  $\mathbf{w}$  such that  $\mathbb{E}_{\mathbf{w}}[d(\mathbf{w}, \mathbf{0})] = D$ . This results again in the simple formula (cf. (76))

$$R(D) \geq h(\mathbf{x}) - \max h(\mathbf{w}).$$

Because the parametric bound (77) is more convenient in most cases and already allows us to separate the source from the distortion, we will concentrate on a generalization of (77) to rectifiable random variables. To this end, we will use the characterization of the RD function in (74) with a specific choice of the function  $\alpha_s$ .

*Theorem 55:* The RD function of an  $m$ -rectifiable random variable  $\mathbf{x}$  on  $\mathbb{R}^M$  with support  $\mathcal{E}$  is lower bounded by

$$R(D) \geq R_{\text{SLB}}(D, s) \triangleq \mathfrak{h}^m(\mathbf{x}) - sD - \log \gamma(s) \quad (79)$$

for each  $s \geq 0$ , where

$$\gamma(s) \triangleq \sup_{\mathbf{y} \in \mathbb{R}^M} \int_{\mathcal{E}} e^{-sd(\mathbf{x}, \mathbf{y})} d\mathcal{H}^m(\mathbf{x}), \quad s \geq 0. \quad (80)$$

<sup>16</sup>Although in [26, Th. 2.3]  $\alpha_s(\mathbf{x}) \geq 1$  is assumed, (74) also holds for  $\alpha_s(\mathbf{x}) > 0$  because of [26, eq. (1.23)].

*Proof:* We start by noting that (80) implies

$$\int_{\mathcal{E}} e^{-sd(\mathbf{x}, \mathbf{y})} d\mathcal{H}^m(\mathbf{x}) \leq \gamma(s) \quad (81)$$

for all  $\mathbf{y} \in \mathbb{R}^M$ . Let  $s \geq 0$  be fixed. By (74),

$$R(D) \geq -sD + \mathbb{E}_{\mathbf{x}}[\log \alpha_s(\mathbf{x})] \quad (82)$$

for every function  $\alpha_s$  satisfying (75). We have (cf. (15))

$$\begin{aligned} \mathbb{E}_{\mathbf{x}} \left[ \frac{1}{\theta_{\mathbf{x}}^m(\mathbf{x})\gamma(s)} e^{-sd(\mathbf{x}, \mathbf{y})} \right] \\ &= \int_{\mathcal{E}} \frac{1}{\theta_{\mathbf{x}}^m(\mathbf{x})\gamma(s)} e^{-sd(\mathbf{x}, \mathbf{y})} \theta_{\mathbf{x}}^m(\mathbf{x}) d\mathcal{H}^m(\mathbf{x}) \\ &= \frac{1}{\gamma(s)} \int_{\mathcal{E}} e^{-sd(\mathbf{x}, \mathbf{y})} d\mathcal{H}^m(\mathbf{x}) \\ &\stackrel{(81)}{\leq} \frac{\gamma(s)}{\gamma(s)} \\ &= 1 \end{aligned}$$

for all  $\mathbf{y} \in \mathbb{R}^M$ . Therefore, the choice  $\alpha_s(\mathbf{x}) \triangleq \frac{1}{\theta_{\mathbf{x}}^m(\mathbf{x})\gamma(s)}$  satisfies (75). Inserting  $\alpha_s(\mathbf{x}) = \frac{1}{\theta_{\mathbf{x}}^m(\mathbf{x})\gamma(s)}$  into (82), we obtain

$$\begin{aligned} R(D) &\geq -sD + \mathbb{E}_{\mathbf{x}} \left[ \log \frac{1}{\theta_{\mathbf{x}}^m(\mathbf{x})\gamma(s)} \right] \\ &= -\mathbb{E}_{\mathbf{x}}[\log \theta_{\mathbf{x}}^m(\mathbf{x})] - sD - \mathbb{E}_{\mathbf{x}}[\log \gamma(s)] \\ &= \mathfrak{h}^m(\mathbf{x}) - sD - \log \gamma(s). \end{aligned}$$

■

For a continuous random variable  $\mathbf{x}$  with positive probability density function almost everywhere (i.e.,  $M$ -rectifiable with support  $\mathbb{R}^M$ ), the definitions of  $\tilde{\gamma}(s)$  in (78) and  $\gamma(s)$  in (80) coincide. Indeed, because  $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{x} - \mathbf{y}, \mathbf{0})$  and a translation of the integrand by  $\mathbf{y}$  does not change the value of the integral over  $\mathbb{R}^M$ , the right-hand side of (78) can be written as (recall that  $\mathcal{H}^M = \mathcal{L}^M$ )

$$\int_{\mathbb{R}^M} e^{-sd(\mathbf{x}, \mathbf{0})} d\mathcal{L}^M(\mathbf{x}) = \int_{\mathbb{R}^M} e^{-sd(\mathbf{x}, \mathbf{y})} d\mathcal{H}^M(\mathbf{x}) \quad (83)$$

for any  $\mathbf{y} \in \mathbb{R}^M$ . Because the left-hand side of (83) does not depend on  $\mathbf{y}$ , taking the supremum over  $\mathbf{y} \in \mathbb{R}^M$  in (83) results in

$$\int_{\mathbb{R}^M} e^{-sd(\mathbf{x}, \mathbf{0})} d\mathcal{L}^M(\mathbf{x}) = \sup_{\mathbf{y} \in \mathbb{R}^M} \int_{\mathbb{R}^M} e^{-sd(\mathbf{x}, \mathbf{y})} d\mathcal{H}^M(\mathbf{x})$$

which is (80). Thus, for a continuous random variable  $\mathbf{x}$  with positive probability density function almost everywhere, the Shannon lower bounds (77) and (79) coincide. However, for a continuous random variable  $\mathbf{x}$  whose support  $\mathcal{E}$  is a proper subset of  $\mathbb{R}^M$  we have  $\gamma(s) \leq \tilde{\gamma}(s)$ , and thus the Shannon lower bound (79) is tighter (i.e., larger) than (77). This is due to the fact that (79) incorporates the additional information that the random variable is restricted to  $\mathcal{E}$ .

## B. Maximizing the Shannon Lower Bound

The optimal choice of  $s$  in (79) depends on  $D$  and is hard to find in general. At least, the following lemma states that the optimal (i.e., largest) lower bound in (79),

$$R_{\text{SLB}}^*(D) \triangleq \sup_{s \geq 0} R_{\text{SLB}}(D, s)$$

is achieved for a finite  $s$ . We recall that  $D_0$  is the infimum of all  $D \geq 0$  such that  $R(D)$  is finite.

*Lemma 56:* Let  $\mathbf{x}$  be an  $m$ -rectifiable random variable with support  $\mathcal{E}$  and finite  $m$ -dimensional entropy  $\mathfrak{h}^m(\mathbf{x})$ . Then for  $D > D_0$  the lower bound  $R_{\text{SLB}}(D, s)$  in (79) satisfies

$$\lim_{s \rightarrow \infty} R_{\text{SLB}}(D, s) = -\infty.$$

*Proof:* See Appendix K. ■

If  $R_{\text{SLB}}(D, s)$  is a continuous function of  $s$ , Lemma 56 implies that for a fixed  $D > D_0$ , the global maximum of  $R_{\text{SLB}}(D, s)$  with respect to  $s$  exists and is either a local maximum or the boundary point  $s = 0$ , i.e.,  $R_{\text{SLB}}^*(D) = R_{\text{SLB}}(D, s)$  for some finite  $s \geq 0$ . Moreover, if  $\gamma(s)$  in (80) is differentiable, we can characterize the local maxima of  $R_{\text{SLB}}(D, s)$  as follows.

*Theorem 57:* Let  $\mathbf{x}$  be an  $m$ -rectifiable random variable with support  $\mathcal{E}$ , and let  $\gamma(s)$  be differentiable. Then for  $D > D_0$ , the lower bound  $R_{\text{SLB}}(D, s)$  in (79) is maximized either for  $s = 0$  or for some  $s > 0$  satisfying  $\tilde{D}(s) = D$ , where

$$\tilde{D}(s) \triangleq -\frac{\gamma'(s)}{\gamma(s)}.$$

That is, the largest lower bound is given by

$$R_{\text{SLB}}^*(D) = \max \left\{ R_{\text{SLB}}(D, 0), \sup_{s > 0: \tilde{D}(s) = D} R_{\text{SLB}}(D, s) \right\}. \quad (84)$$

*Proof:* We recall from (79) that  $R_{\text{SLB}}(D, s) = \mathfrak{h}^m(\mathbf{x}) - sD - \log \gamma(s)$ . Thus, because  $\gamma(s)$  is differentiable, a necessary condition for a local maximum of  $R_{\text{SLB}}(D, s)$  with respect to  $s$  is obtained by setting to zero the derivative of  $R_{\text{SLB}}(D, s)$  with respect to  $s$ . Solving the resulting equation for  $D$  yields  $\tilde{D}(s) = D$ . Thus, for a given  $D > D_0$ ,  $R_{\text{SLB}}(D, s)$  can only have a local maximum at  $s \in (0, \infty)$  satisfying  $\tilde{D}(s) = D$ . By Lemma 56, the global maximum either is a local maximum or is achieved for  $s = 0$ , which concludes the proof. ■

If  $\gamma(s)$  is differentiable, Theorem 57 provides a “parametrization” of the graph of the largest bound  $R_{\text{SLB}}^*(D)$ , i.e., we can characterize the set

$$\mathcal{G} \triangleq \{(D, R_{\text{SLB}}^*(D)) \in \mathbb{R}^2 : D > D_0\}. \quad (85)$$

As a basis for this characterization, we define the sets

$$\begin{aligned} \mathcal{F}_1 &\triangleq \{(\tilde{D}(s), R_{\text{SLB}}(\tilde{D}(s), s)) : s > 0\} \\ \mathcal{F}_2 &\triangleq \{(D, \mathfrak{h}^m(\mathbf{x}) - \log \mathcal{H}^m(\mathcal{E})) : D > D_0\} \end{aligned} \quad (86)$$

which are illustrated in Fig. 1. Note that  $\mathcal{F}_1$  is not necessarily the graph of a function, whereas  $\mathcal{F}_2$  constitutes a horizontal line in the  $(D, R)$  plane.

*Corollary 58:* Let  $\mathbf{x}$  be an  $m$ -rectifiable random variable with support  $\mathcal{E}$ , and let  $\gamma(s)$  be differentiable. Define  $\mathcal{F} \triangleq$

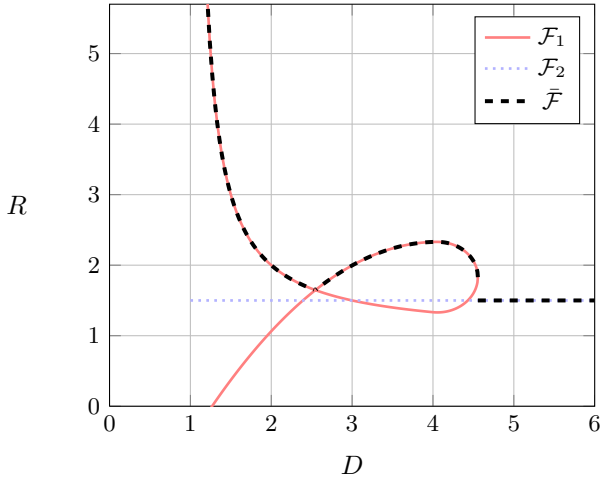


Fig. 1. Illustration of the sets  $\mathcal{F}_1$ ,  $\mathcal{F}_2$ , and  $\tilde{\mathcal{F}}$  (assuming  $D_0 = 1$ ).

$\mathcal{F}_1 \cup \mathcal{F}_2$ . Then  $\mathcal{G} = \tilde{\mathcal{F}}$ , where  $\tilde{\mathcal{F}}$  is the upper envelope of  $\mathcal{F}$  given by

$$\tilde{\mathcal{F}} \triangleq \left\{ (D, R) \in \mathcal{F} : R = \max_{(D, R') \in \mathcal{F}} R' \right\}. \quad (87)$$

*Proof:* All elements  $(D, R) \in \mathcal{F}$  can be written as  $(D, R) = (D, R_{\text{SLB}}(D, s))$  for some  $s \geq 0$ . Indeed, for  $(D, R) \in \mathcal{F}_1$  this is obvious, and for  $(D, R) \in \mathcal{F}_2$  we have

$$R = \mathfrak{h}^m(\mathbf{x}) - \log \mathcal{H}^m(\mathcal{E}) \stackrel{(a)}{=} \mathfrak{h}^m(\mathbf{x}) - \log \gamma(0) \stackrel{(79)}{=} R_{\text{SLB}}(D, 0) \quad (88)$$

where (a) holds because  $\gamma(0) \stackrel{(80)}{=} \int_{\mathcal{E}} 1 \, d\mathcal{H}^m(\mathbf{x}) = \mathcal{H}^m(\mathcal{E})$ . Hence, for all  $(D, R) \in \mathcal{F}$ , we obtain

$$R \leq \sup_{s \geq 0} R_{\text{SLB}}(D, s) = R_{\text{SLB}}^*(D). \quad (89)$$

Because  $\tilde{\mathcal{F}} \subseteq \mathcal{F}$ , (89) also holds for  $(D, R) \in \tilde{\mathcal{F}}$ .

Consider now the pair  $(D, R) \in \tilde{\mathcal{F}}$  for a fixed  $D > D_0$ . By (87), for a pair  $(D, R') \in \mathcal{F}$  we obtain  $R \geq R'$ . In particular, for  $s > 0$  satisfying  $\tilde{D}(s) = D$ , the pair  $(D, R_{\text{SLB}}(D, s))$  belongs to  $\mathcal{F}_1 \subseteq \mathcal{F}$ , and thus

$$R \geq R_{\text{SLB}}(D, s). \quad (90)$$

Similarly,  $(D, \mathfrak{h}^m(\mathbf{x}) - \log \mathcal{H}^m(\mathcal{E})) \in \mathcal{F}_2 \subseteq \mathcal{F}$ , and thus

$$R \geq \mathfrak{h}^m(\mathbf{x}) - \log \mathcal{H}^m(\mathcal{E}) \stackrel{(88)}{=} R_{\text{SLB}}(D, 0). \quad (91)$$

Combining (90) for all  $s > 0$  satisfying  $\tilde{D}(s) = D$  and (91), we obtain

$$R \geq \max \left\{ R_{\text{SLB}}(D, 0), \sup_{s > 0: \tilde{D}(s) = D} R_{\text{SLB}}(D, s) \right\} \stackrel{(84)}{=} R_{\text{SLB}}^*(D). \quad (92)$$

Combining (89) and (92) for an arbitrary  $(D, R) \in \tilde{\mathcal{F}}$  implies that  $R = R_{\text{SLB}}^*(D)$ . By (85), this yields  $(D, R) \in \mathcal{G}$  and thus  $\tilde{\mathcal{F}} \subseteq \mathcal{G}$ . Because both sets  $\mathcal{G}$  and  $\tilde{\mathcal{F}}$  contain exactly one element  $(D, R)$  for each  $D > D_0$ , we obtain  $\tilde{\mathcal{F}} = \mathcal{G}$ . ■

In certain cases, it may not be possible to differentiate  $\gamma(s)$ , and thus the direct calculation of  $\tilde{D}(s) = -\gamma'(s)/\gamma(s)$  is not possible. However, one can show that, under certain

smoothness conditions, the supremum in (80) is in fact a maximum, i.e.,

$$\gamma(s) = \max_{\mathbf{y} \in \mathbb{R}^M} \int_{\mathcal{E}} e^{-sd(\mathbf{x}, \mathbf{y})} \, d\mathcal{H}^m(\mathbf{x}) \quad (93)$$

and  $\tilde{D}(s)$  can be rewritten as

$$\tilde{D}(s) = D^*(s) \triangleq \frac{1}{\gamma(s)} \int_{\mathcal{E}} d(\mathbf{x}, \tilde{\mathbf{y}}(s)) e^{-sd(\mathbf{x}, \tilde{\mathbf{y}}(s))} \, d\mathcal{H}^m(\mathbf{x})$$

where  $\tilde{\mathbf{y}}(s)$  is the maximizing value in the definition of  $\gamma(s)$  (cf. (80)):

$$\tilde{\mathbf{y}}(s) \triangleq \arg \max_{\mathbf{y} \in \mathbb{R}^M} \int_{\mathcal{E}} e^{-sd(\mathbf{x}, \mathbf{y})} \, d\mathcal{H}^m(\mathbf{x}).$$

(Thus,  $\gamma(s) = \int_{\mathcal{E}} e^{-sd(\mathbf{x}, \tilde{\mathbf{y}}(s))} \, d\mathcal{H}^m(\mathbf{x})$ .) The following corollary shows that even if we do not know whether  $\gamma(s)$  is differentiable, we can construct a set  $\tilde{\mathcal{F}}$  of lower bounds on the RD function. To this end, we define  $\tilde{\mathcal{F}} \triangleq \tilde{\mathcal{F}}_1 \cup \mathcal{F}_2$ , where

$$\tilde{\mathcal{F}}_1 \triangleq \{(D^*(s), R_{\text{SLB}}(D^*(s), s)) : s > 0\}$$

and  $\mathcal{F}_2$  was defined in (86).

*Corollary 59:* Let  $\mathbf{x}$  be an  $m$ -rectifiable random variable with support  $\mathcal{E}$ . Then  $\tilde{\mathcal{F}}$  is a set of lower bounds on the RD function, i.e., for each  $(D, R) \in \tilde{\mathcal{F}}$ , we have  $R(D) \geq R$ .

*Proof:* Let  $(D, R) \in \tilde{\mathcal{F}}$ .

*Case  $(D, R) \in \tilde{\mathcal{F}}_1$ :* In this case, we have  $(D, R) = (D^*(s), R_{\text{SLB}}(D^*(s), s))$  for some  $s > 0$ . Thus,  $R = R_{\text{SLB}}(D^*(s), s) = R_{\text{SLB}}(D, s)$  and, by (79),  $R \leq R(D)$ .

*Case  $(D, R) \in \mathcal{F}_2$ :* In this case, as in (88), we have  $R = R_{\text{SLB}}(D, 0)$ . By (79), we have  $R_{\text{SLB}}(D, 0) \leq R(D)$ , which implies  $R \leq R(D)$ .

In either case  $R \leq R(D)$ , which concludes the proof. ■

By Corollary 59, we can use the sets  $\tilde{\mathcal{F}}_1$  and  $\mathcal{F}_2$  to construct lower bounds on the RD function.<sup>17</sup> More specifically, these bounds are obtained via the following program:

- (P1) Calculate  $D^*(s)$  for  $s \in (0, \infty)$ .
- (P2) Plot the  $s$ -parametrized curve  $(D^*(s), R_{\text{SLB}}(D^*(s), s))$  for  $s \in (0, \infty)$ .
- (P3) Plot the horizontal line  $(D, \mathfrak{h}^m(\mathbf{x}) - \log \mathcal{H}^m(\mathcal{E}))$  for  $D \in (D_0, \infty)$ .
- (P4) Take the upper envelope of these two curves.

In the subsequent Section X-C, we will apply the program (P1)–(P4) to a specific example.

### C. Shannon Lower Bound on the Unit Circle

To demonstrate the practical relevance of Theorem 55, we apply it to the simple example given by  $\mathcal{E} = \mathcal{S}_1$ , i.e., the unit circle in  $\mathbb{R}^2$ , and squared error distortion, i.e.,  $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^2$ . In order to calculate  $\gamma(s)$ , we first show that it can be expressed as in (93), i.e.,

$$\gamma(s) = \max_{\mathbf{y} \in \mathbb{R}^2} \int_{\mathcal{S}_1} e^{-s\|\mathbf{x} - \mathbf{y}\|^2} \, d\mathcal{H}^1(\mathbf{x})$$

<sup>17</sup>If  $\tilde{\mathcal{F}}_1 = \mathcal{F}_1$ , we obtain by Corollary 58 that these bounds will be the best Shannon lower bounds. However, explicit smoothness conditions that guarantee  $\tilde{\mathcal{F}}_1 = \mathcal{F}_1$  are difficult to find.

for all  $s \geq 0$ . Let  $s \geq 0$  be arbitrary but fixed. Note that we can restrict to  $\mathbf{y} = (y_1 \ 0)^T$ , with  $y_1 \geq 0$ , because the problem is invariant under rotations. Thus,

$$\int_{S_1} e^{-s\|\mathbf{x}-\mathbf{y}\|^2} d\mathcal{H}^1(\mathbf{x}) = \int_{S_1} e^{-s((x_1-y_1)^2+x_2^2)} d\mathcal{H}^1(\mathbf{x})$$

and therefore we have to maximize the function

$$f_s(y_1) \triangleq \int_{S_1} e^{-s((x_1-y_1)^2+x_2^2)} d\mathcal{H}^1(\mathbf{x})$$

on  $[0, \infty)$ . To this end, we consider the derivative  $f'_s$  and change the order of differentiation and integration (according to [27, Cor. 5.9], this is justified because  $\mathcal{H}^1|_{S_1}$  is a finite measure and  $0 < e^{-s((x_1-y_1)^2+x_2^2)} \leq 1$  for  $(x_1 \ x_2)^T \in S_1$ ). This results in the expression

$$f'_s(y_1) = \int_{S_1} 2s(x_1 - y_1) e^{-s((x_1-y_1)^2+x_2^2)} d\mathcal{H}^1(\mathbf{x}). \quad (94)$$

Because  $x_1 \leq 1$  for  $\mathbf{x} \in S_1$ , we have  $f'_s(y_1) < 0$  for  $y_1 > 1$ , i.e.,  $f_s$  is monotonically decreasing on  $(1, \infty)$ . Thus, the function  $f_s$  can only attain its maximum in the compact interval  $[0, 1]$ . Because  $f_s$  is a continuous function, we conclude that  $\gamma(s) = \max_{\mathbf{y} \in \mathbb{R}^2} \int_{S_1} e^{-s\|\mathbf{x}-\mathbf{y}\|^2} d\mathcal{H}^1(\mathbf{x})$  exists for each  $s \geq 0$ .

To characterize  $\gamma(s)$  in more detail, we consider the equation  $f'_s(y_1) = 0$  to find local maxima. By (94) and because  $x_1^2 + x_2^2 = 1$  for  $\mathbf{x} \in S_1$ ,  $f'_s(y_1) = 0$  is equivalent to

$$2se^{-s(1+y_1^2)} \int_{S_1} (x_1 - y_1) e^{2sx_1y_1} d\mathcal{H}^1(\mathbf{x}) = 0. \quad (95)$$

Furthermore, because  $2se^{-s(1+y_1^2)} > 0$  and using the transformation  $x_1 = \cos \phi$ ,  $x_2 = \sin \phi$ , we obtain that (95) is equivalent to

$$\int_0^{2\pi} (\cos \phi - y_1) e^{2sy_1 \cos \phi} d\phi = 0. \quad (96)$$

Because we know that the function  $f'_s$  can only have zeros on  $[0, 1]$ , we can solve (96) numerically for any fixed  $s \geq 0$  and compare the values  $f_s(y_1)$  at the different solutions  $y_1$  and at the boundary points  $y_1 = 0$  and  $y_1 = 1$  to find  $\gamma(s)$ . In Fig. 2, the values of  $\gamma(s)$  are depicted for  $s \in [0.01, 5000]$ .

We now have all the ingredients to calculate the parametric lower bound  $R_{\text{SLB}}(D, s)$  in (79) for any given distortion  $D$  and an arbitrary source  $\mathbf{x}$  on  $S_1$ . In particular, let us consider a uniform distribution of  $\mathbf{x}$  on  $S_1$ , where  $h^1(\mathbf{x}) = \log(2\pi)$  (see (37)). In Fig. 3, we show the lower bound  $R_{\text{SLB}}(D, s)$  for  $s \in [1, 94]$  and distortion  $D = 10^{-2}$ . It can be seen that the maximal lower bound  $R_{\text{SLB}}(10^{-2}, s)$  is obtained for  $s \approx 50$ .

To plot Fig. 3, we had to calculate  $\gamma(s)$  for many different values of  $s$ . We also used “trial and error” to find the region of  $s$  where the maximal lower bound  $R_{\text{SLB}}(10^{-2}, s)$  arises. To avoid this tedious optimization procedure, which would have to be carried out for each value of  $D$  under consideration, we can use the program (P1)–(P4) formulated in Section X-B. In Fig. 4, we show the lower bounds on  $R(D)$  resulting from this program for  $s \in [1, 10^5]$ , which corresponds to  $D \in [5 \cdot 10^{-5}, 1]$ . We also show in Fig. 4 an upper bound on  $R(D)$  using the following result.

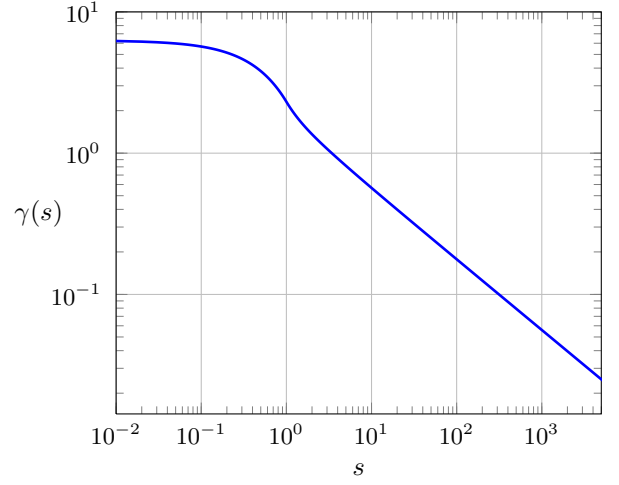


Fig. 2. Graph of  $\gamma(s)$  for  $s \in [0.01, 5000]$ .

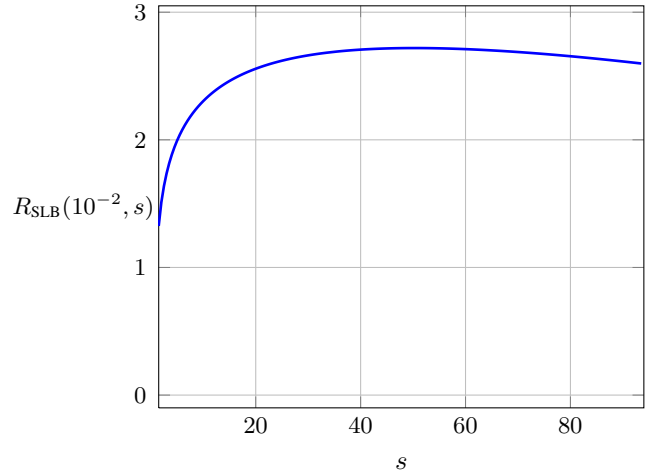


Fig. 3. Shannon lower bound  $R_{\text{SLB}}(10^{-2}, s)$  for  $s \in [1, 94]$ .

**Theorem 60:** Let the random variable  $\mathbf{x}$  on  $\mathbb{R}^2$  be uniformly distributed on the unit circle, and consider squared error distortion, i.e.,  $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^2$ . For any  $n \in \mathbb{N}$ ,

$$R(\bar{D}_n) \leq \log n \quad (97)$$

where

$$\bar{D}_n = 1 - \left( \frac{n}{\pi} \sin \frac{\pi}{n} \right)^2. \quad (98)$$

*Proof:* See Appendix L. ■

The upper bound depicted in Fig. 4 was obtained by linearly interpolating the upper bounds (97) corresponding to different values of  $n$  (and, hence, of  $\bar{D}_n$ ). This is justified by the convexity of the RD function [13, Lem. 10.4.1]. Note that the lower and upper bounds shown in Fig. 4 are quite close, and thus they provide a rather accurate characterization of the RD function of  $\mathbf{x}$ .

## XI. CONCLUSION

We presented a generalization of entropy to singular random variables supported on integer-dimensional subsets of Euclidean space. More specifically, we considered random

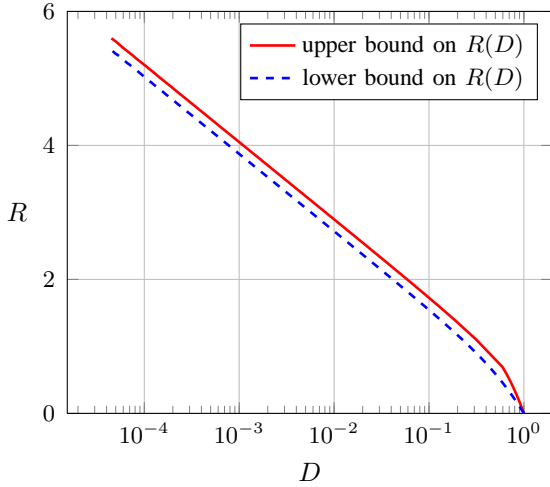


Fig. 4. Shannon lower bound constructed by (P1)–(P4) and upper bound (97) for a source  $\mathbf{x}$  on  $\mathbb{R}^2$  uniformly distributed on the unit circle and squared error distortion.

variables distributed according to a rectifiable measure. Similar to continuous random variables, these rectifiable random variables can be described by a density. However, in contrast to continuous random variables, the density is nonzero only on a lower-dimensional subset and has to be integrated with respect to a Hausdorff measure to calculate probabilities. Our entropy definition is based on this Hausdorff density but otherwise resembles the usual definition of differential entropy. However, this formal similarity has to be interpreted with caution because Hausdorff measures and projections of rectifiable sets do not always conform to intuition. We thus emphasized mathematical rigor and carefully stated all the assumptions underlying our results.

We showed that for the special cases of rectifiable random variables given by discrete and continuous random variables, our entropy definition reduces to classical entropy and differential entropy, respectively. Furthermore, we established a connection between our entropy and differential entropy for a rectifiable random variable that is obtained from a continuous random variable through a one-to-one transformation. For joint and conditional entropy, our analysis showed that the geometry of the support sets of the random variables plays an important role. This role is evidenced by the facts that the chain rule may contain a geometric correction term and conditioning may increase entropy.

Random variables that are neither discrete nor continuous are not only of theoretical interest. Continuity of a random variable cannot be assumed if there are deterministic dependencies reducing the intrinsic dimension of the random variable, which is especially likely to occur in high-dimensional problems. As two basic examples, we considered a random variable  $\mathbf{x} \in \mathbb{R}^2$  supported on the unit circle, which is intrinsically only one-dimensional, and the class of positive semidefinite rank-one random matrices. In both cases, the differential entropy is not defined and, in fact, classical information theory does not provide a rigorous definition of entropy for these random variables.

As an application of our entropy definition to source coding,

we provided a characterization of the minimal codeword length for quantizations of integer-dimensional sources. Furthermore, we presented a result in rate-distortion theory that generalizes the Shannon lower bound for discrete and continuous random variables to the larger class of rectifiable random variables. The usefulness of this bound was demonstrated by the example of a uniform source on the unit circle. The resulting bound appears to be the first rigorous lower bound on the rate-distortion function for that distribution.

Possible directions for future work include the extension of our entropy definition to distributions mixing different dimensions (e.g., discrete-continuous mixtures). The extension to noninteger-dimensional singular distributions seems to be possible only in terms of upper and lower entropies, which could be defined based on the upper and lower Hausdorff densities<sup>18</sup> [14, Def. 2.55]. Furthermore, our entropy can be extended to infinite-length sequences of rectifiable random variables, which leads to the definition of an entropy rate generalizing the (differential) entropy rate of a sequence of discrete or continuous random variables. Finally, applications of our entropy to source coding and channel coding problems involving integer-dimensional singular random variables are largely unexplored.

#### APPENDIX A PROOF OF LEMMA 9

To prove the existence of a support  $\tilde{\mathcal{E}} \subseteq \mathcal{E}$ , we have to construct a set  $\tilde{\mathcal{E}}$  that satisfies (cf. Definition 8)

- (i)  $\tilde{\mathcal{E}} = \bigcup_{k \in \mathbb{N}} f_k(\mathcal{C}_k)$  where, for  $k \in \mathbb{N}$ ,  $\mathcal{C}_k \subseteq \mathbb{R}^m$  is a bounded Borel set and  $f_k: \mathbb{R}^m \rightarrow \mathbb{R}^M$  is a Lipschitz function that is one-to-one on  $\mathcal{C}_k$ ;
- (ii)  $\tilde{\mathcal{E}} \subseteq \mathcal{E}$ ;
- (iii)  $\mu \ll \mathcal{H}^m|_{\tilde{\mathcal{E}}}$ ;
- (iv)  $\frac{d\mu}{d\mathcal{H}^m|_{\tilde{\mathcal{E}}}} > 0$   $\mathcal{H}^m|_{\tilde{\mathcal{E}}}$ -almost everywhere.

To prove (i), we note that, by (9), the  $m$ -rectifiable set  $\mathcal{E}$  satisfies  $\mathcal{E} \subseteq \mathcal{E}_0 \cup \bigcup_{k \in \mathbb{N}} f_k(\mathcal{A}_k)$  with bounded Borel sets  $\mathcal{A}_k \subseteq \mathbb{R}^m$ , Lipschitz functions  $f_k: \mathbb{R}^m \rightarrow \mathbb{R}^M$  that are one-to-one on  $\mathcal{A}_k$ , and  $\mathcal{H}^m(\mathcal{E}_0) = 0$ . Because  $\mu \ll \mathcal{H}^m|_{\mathcal{E}}$ , we obtain  $\mu \ll \mathcal{H}^m|_{\mathcal{E}^*}$  where  $\mathcal{E}^* \triangleq \bigcup_{k \in \mathbb{N}} f_k(\mathcal{A}_k)$ . Thus, the Radon-Nikodym derivative  $\frac{d\mu}{d\mathcal{H}^m|_{\mathcal{E}^*}}$  exists. Note that  $\frac{d\mu}{d\mathcal{H}^m|_{\mathcal{E}^*}}$  is in fact an equivalence class of measurable functions and only defined up to a set of  $\mathcal{H}^m|_{\mathcal{E}^*}$ -measure zero. Because  $\mu(\mathcal{E}^c) = 0$  and  $\mu((\mathcal{E}^*)^c) = 0$ , we can choose a function  $g$  in the equivalence class of  $\frac{d\mu}{d\mathcal{H}^m|_{\mathcal{E}^*}}$  satisfying  $g(x) = 0$  on  $(\mathcal{E} \cap \mathcal{E}^*)^c$ . Since  $g$  is a measurable function, the set  $g^{-1}(\{0\})$  is  $\mathcal{H}^m$ -measurable. Furthermore, because  $\mathcal{E}^*$  is  $m$ -rectifiable, Property 1 in Lemma 4 implies that the subset  $g^{-1}(\{0\}) \cap \mathcal{E}^*$  is again  $m$ -rectifiable. By [28, Lem. 15.5(4)], there exists a Borel set  $\mathcal{B}_0$  satisfying

$$\mathcal{B}_0 \supseteq g^{-1}(\{0\}) \cap \mathcal{E}^* \quad (99)$$

and  $\mathcal{H}^m(\mathcal{B}_0 \setminus (g^{-1}(\{0\}) \cap \mathcal{E}^*)) = 0$ . The absolute continuity  $\mu \ll \mathcal{H}^m|_{\mathcal{E}^*} \ll \mathcal{H}^m$  then implies

$$\mu(\mathcal{B}_0 \setminus (g^{-1}(\{0\}) \cap \mathcal{E}^*)) = 0. \quad (100)$$

<sup>18</sup>The upper and lower Hausdorff densities exist for arbitrary distributions, whereas, by Preiss' Theorem [16, Th. 5.6], the existence of the Hausdorff density implies that the measure is rectifiable.

We further have

$$\begin{aligned} \mu(\mathcal{B}_0) &\leq \mu(\mathcal{B}_0 \setminus (g^{-1}(\{0\}) \cap \mathcal{E}^*)) + \mu(g^{-1}(\{0\}) \cap \mathcal{E}^*) \\ &\leq \mu(\mathcal{B}_0 \setminus (g^{-1}(\{0\}) \cap \mathcal{E}^*)) + \mu(g^{-1}(\{0\})) \\ &\stackrel{(a)}{=} 0 \end{aligned} \quad (101)$$

where (a) holds by (100) and because  $\mu(g^{-1}(\{0\})) = \int_{g^{-1}(\{0\})} g(\mathbf{x}) d\mathcal{H}^m|_{\mathcal{E}^*}(\mathbf{x}) = 0$ . Let us define

$$\tilde{\mathcal{E}} \triangleq \bigcup_{k \in \mathbb{N}} f_k(\mathcal{A}_k \setminus f_k^{-1}(\mathcal{B}_0)) \quad (102)$$

where  $\mathcal{A}_k \setminus f_k^{-1}(\mathcal{B}_0)$  are bounded Borel sets (this is because  $\mathcal{A}_k$  are bounded Borel sets,  $f_k$  are continuous functions, and  $\mathcal{B}_0$  is a Borel set). As  $f_k$  are Lipschitz functions that are one-to-one on  $\mathcal{A}_k$ , and thus also on  $\mathcal{A}_k \setminus f_k^{-1}(\mathcal{B}_0)$ , this shows (i).

Next, we prove (ii). We have  $\mathbf{y} \in f_k(\mathcal{A}_k \setminus f_k^{-1}(\mathcal{B}_0))$  if and only if there exists  $\mathbf{x} \in \mathcal{A}_k \setminus f_k^{-1}(\mathcal{B}_0)$  such that  $f_k(\mathbf{x}) = \mathbf{y}$ , which in turn holds if and only if there exists  $\mathbf{x}' \in \mathcal{A}_k$  such that  $f_k(\mathbf{x}') = \mathbf{y}$  and  $\mathbf{y} \notin \mathcal{B}_0$ . Hence,  $f_k(\mathcal{A}_k \setminus f_k^{-1}(\mathcal{B}_0)) = f_k(\mathcal{A}_k) \setminus \mathcal{B}_0$ . We can thus rewrite  $\tilde{\mathcal{E}}$  in (102) as

$$\tilde{\mathcal{E}} = \bigcup_{k \in \mathbb{N}} f_k(\mathcal{A}_k) \setminus \mathcal{B}_0 = \mathcal{E}^* \setminus \mathcal{B}_0 \subseteq \mathcal{E}^* \setminus (g^{-1}(\{0\}) \cap \mathcal{E}^*) \quad (103)$$

where the final inclusion holds by (99). Because we chose  $g(\mathbf{x}) = 0$  on  $(\mathcal{E} \cap \mathcal{E}^*)^c = \mathcal{E}^c \cup (\mathcal{E}^*)^c$ , we obtain  $\mathcal{E}^c \subseteq g^{-1}(\{0\})$ . Inserting this into (103) yields

$$\tilde{\mathcal{E}} \subseteq \mathcal{E}^* \setminus (\mathcal{E}^c \cap \mathcal{E}^*) = \mathcal{E}^* \cap (\mathcal{E} \cup (\mathcal{E}^*)^c) = \mathcal{E}^* \cap \mathcal{E} \subseteq \mathcal{E}$$

which is (ii).

To prove (iii), we start with an arbitrary  $\mathcal{H}^m$ -measurable set  $\mathcal{A} \subseteq \mathbb{R}^M$  with  $\mathcal{H}^m|_{\mathcal{E}^*}(\mathcal{A}) = 0$ . We have

$$\begin{aligned} \mathcal{H}^m|_{\mathcal{E}^*}(\mathcal{A} \setminus \mathcal{B}_0) &= \mathcal{H}^m(\mathcal{E}^* \cap (\mathcal{A} \setminus \mathcal{B}_0)) \\ &= \mathcal{H}^m(\mathcal{E}^* \cap \mathcal{A} \cap \mathcal{B}_0^c) \\ &= \mathcal{H}^m((\mathcal{E}^* \setminus \mathcal{B}_0) \cap \mathcal{A}) \\ &\stackrel{(a)}{=} \mathcal{H}^m(\tilde{\mathcal{E}} \cap \mathcal{A}) \\ &= \mathcal{H}^m|_{\tilde{\mathcal{E}}}(\mathcal{A}) \\ &= 0 \end{aligned}$$

where (a) holds because  $\tilde{\mathcal{E}} = \mathcal{E}^* \setminus \mathcal{B}_0$  by (103). Because  $\mu \ll \mathcal{H}^m|_{\mathcal{E}^*}$ , this implies  $\mu(\mathcal{A} \setminus \mathcal{B}_0) = 0$  and, since  $\mu(\mathcal{B}_0) = 0$  by (101), we obtain  $\mu(\mathcal{A}) = 0$ . Thus,  $\mathcal{H}^m|_{\tilde{\mathcal{E}}}(\mathcal{A}) = 0$  implies  $\mu(\mathcal{A}) = 0$ , which proves (iii).

To prove (iv), we first show that  $g$  is also in the equivalence class of the Radon-Nikodym derivative  $\frac{d\mu}{d\mathcal{H}^m|_{\mathcal{E}}}$ . Indeed, we

have for an arbitrary measurable set  $\mathcal{A} \subseteq \mathbb{R}^M$

$$\begin{aligned} \mu(\mathcal{A}) &= \int_{\mathcal{A}} g(\mathbf{x}) d\mathcal{H}^m|_{\mathcal{E}^*}(\mathbf{x}) \\ &= \int_{\mathcal{A} \cap \tilde{\mathcal{E}}} g(\mathbf{x}) d\mathcal{H}^m|_{\mathcal{E}^*}(\mathbf{x}) + \int_{\mathcal{A} \cap \tilde{\mathcal{E}}^c} g(\mathbf{x}) d\mathcal{H}^m|_{\mathcal{E}^*}(\mathbf{x}) \\ &\stackrel{(a)}{=} \int_{\mathcal{A}} g(\mathbf{x}) d\mathcal{H}^m|_{\tilde{\mathcal{E}}}(\mathbf{x}) + \int_{\mathcal{A} \cap \tilde{\mathcal{E}}^c} g(\mathbf{x}) d\mathcal{H}^m|_{\mathcal{E}^*}(\mathbf{x}) \\ &= \int_{\mathcal{A}} g(\mathbf{x}) d\mathcal{H}^m|_{\tilde{\mathcal{E}}}(\mathbf{x}) + \mu(\mathcal{A} \cap \tilde{\mathcal{E}}^c) \\ &\stackrel{(b)}{=} \int_{\mathcal{A}} g(\mathbf{x}) d\mathcal{H}^m|_{\tilde{\mathcal{E}}}(\mathbf{x}) \end{aligned}$$

where (a) holds because  $\tilde{\mathcal{E}} \subseteq \mathcal{E}^*$  (see (103)) and (b) holds because  $\mu(\mathcal{A} \cap \tilde{\mathcal{E}}^c) = 0$  (indeed  $\mathcal{H}^m|_{\tilde{\mathcal{E}}}(\mathcal{A} \cap \tilde{\mathcal{E}}^c) = \mathcal{H}^m(\tilde{\mathcal{E}} \cap \mathcal{A} \cap \tilde{\mathcal{E}}^c) = 0$  implies  $\mu(\mathcal{A} \cap \tilde{\mathcal{E}}^c) = 0$  by (iii)). By (103), we have  $\tilde{\mathcal{E}} \subseteq \mathcal{E}^*$ , which implies

$$\mathcal{H}^m|_{\tilde{\mathcal{E}}}((\mathcal{E}^*)^c) = \mathcal{H}^m(\tilde{\mathcal{E}} \cap (\mathcal{E}^*)^c) \leq \mathcal{H}^m(\mathcal{E}^* \cap (\mathcal{E}^*)^c) = 0. \quad (104)$$

By (99), we have  $\mathcal{B}_0^c \subseteq (g^{-1}(\{0\}))^c \cup (\mathcal{E}^*)^c$ . Hence, for  $\mathbf{x} \in \mathcal{B}_0^c$  we have either  $\mathbf{x} \in (g^{-1}(\{0\}))^c$ —which is equivalent to  $g(\mathbf{x}) > 0$ —or  $\mathbf{x} \in (\mathcal{E}^*)^c$ . By (104), we therefore have for  $\mathcal{H}^m|_{\tilde{\mathcal{E}}}$ -almost all  $\mathbf{x} \in \mathcal{B}_0^c$  that  $g(\mathbf{x}) > 0$ . In particular, because, by (103),  $\tilde{\mathcal{E}} = \mathcal{E}^* \setminus \mathcal{B}_0 \subseteq \mathcal{B}_0^c$ , we obtain  $g(\mathbf{x}) > 0$  for  $\mathcal{H}^m|_{\tilde{\mathcal{E}}}$ -almost all  $\mathbf{x} \in \tilde{\mathcal{E}}$ . This proves (iv).

Finally, we show that the support is unique up to sets of  $\mathcal{H}^m$ -measure zero. Let  $\mathcal{E}_1$  and  $\mathcal{E}_2$  be two support sets of an  $m$ -rectifiable measure  $\mu$ , and denote the Radon-Nikodym derivative  $\frac{d\mu}{d\mathcal{H}^m|_{\mathcal{E}_2}}$  by  $g_2$ . Then

$$\int_{\mathcal{E}_2 \setminus \mathcal{E}_1} g_2(\mathbf{x}) d\mathcal{H}^m|_{\mathcal{E}_2}(\mathbf{x}) = \mu(\mathcal{E}_2 \setminus \mathcal{E}_1) = 0 \quad (105)$$

where the latter equality holds because  $\mu(\mathcal{E}_1^c) = 0$  (indeed,  $\mathcal{H}^m|_{\mathcal{E}_1}(\mathcal{E}_1^c) = 0$  implies  $\mu(\mathcal{E}_1^c) = 0$  due to  $\mu \ll \mathcal{H}^m|_{\mathcal{E}_1}$ ). Since by Definition 8  $g_2 > 0$  on  $\mathcal{E}_2$ ,  $\mathcal{H}^m|_{\mathcal{E}_2}$ -almost everywhere, (105) implies  $\mathcal{H}^m(\mathcal{E}_2 \setminus \mathcal{E}_1) = 0$ . By an analogous argument, we obtain  $\mathcal{H}^m(\mathcal{E}_1 \setminus \mathcal{E}_2) = 0$ . This shows that  $\mathcal{E}_1$  and  $\mathcal{E}_2$  coincide up to a set of  $\mathcal{H}^m$ -measure zero.

## APPENDIX B PROOF OF THEOREM 13

*Proof of Part 1:* Let  $\mathbf{x}$  be 0-rectifiable with support  $\mathcal{E}$ , i.e.,  $\mu\mathbf{x}^{-1} \ll \mathcal{H}^0|_{\mathcal{E}}$  for a 0-rectifiable set  $\mathcal{E}$ . Recall that a 0-rectifiable set  $\mathcal{E}$  is by definition countable, i.e.,  $\mathcal{E} = \{\mathbf{x}_i : i \in \mathcal{I}\}$  for a countable index set  $\mathcal{I}$ . By (16),  $\Pr\{\mathbf{x} \in \mathcal{E}\} = 1$ , which implies that  $\mathbf{x}$  is a discrete random variable. Finally,

$$\begin{aligned} p_{\mathbf{x}}(\mathbf{x}_i) &= \Pr\{\mathbf{x} = \mathbf{x}_i\} \\ &\stackrel{(12)}{=} \mu\mathbf{x}^{-1}(\{\mathbf{x}_i\}) \\ &= \int_{\{\mathbf{x}_i\}} \frac{d\mu\mathbf{x}^{-1}}{d\mathcal{H}^0|_{\mathcal{E}}}(\mathbf{x}) d\mathcal{H}^0|_{\mathcal{E}}(\mathbf{x}) \\ &\stackrel{(a)}{=} \frac{d\mu\mathbf{x}^{-1}}{d\mathcal{H}^0|_{\mathcal{E}}}(\mathbf{x}_i) \\ &\stackrel{(13)}{=} \theta_{\mathbf{x}}^0(\mathbf{x}_i) \end{aligned}$$



where (a) holds because  $\mathcal{H}^0$  is the counting measure.

Conversely, let  $\mathbf{x}$  be a discrete random variable taking on the values  $\mathbf{x}_i$ ,  $i \in \mathcal{I}$ . We set  $\mathcal{E} \triangleq \{\mathbf{x}_i : i \in \mathcal{I}\}$ , which is countable and, thus, 0-rectifiable. Because  $\mathcal{E}$  includes all possible values of  $\mathbf{x}$ , we have  $\Pr\{\mathbf{x} \in \mathcal{E}^c\} = \mu\mathbf{x}^{-1}(\mathcal{E}^c) = 0$ . For  $\mathcal{A} \subseteq \mathbb{R}^M$ , the measure  $\mathcal{H}^0|_{\mathcal{E}}(\mathcal{A})$  counts the number of points in  $\mathcal{A}$  that also belong to  $\mathcal{E}$ . Thus, for any set  $\mathcal{A}$  such that  $\mathcal{H}^0|_{\mathcal{E}}(\mathcal{A}) = 0$ , we obtain that  $\mathcal{A} \cap \mathcal{E} = \emptyset$  and hence  $\mathcal{A} \subseteq \mathcal{E}^c$ . This implies  $\mu\mathbf{x}^{-1}(\mathcal{A}) \leq \mu\mathbf{x}^{-1}(\mathcal{E}^c) = 0$ . Thus, we showed that  $\mu\mathbf{x}^{-1}(\mathcal{A}) = 0$  for any set  $\mathcal{A}$  with  $\mathcal{H}^0|_{\mathcal{E}}(\mathcal{A}) = 0$ , i.e.,  $\mu\mathbf{x}^{-1} \ll \mathcal{H}^0|_{\mathcal{E}}$ . Hence,  $\mathbf{x}$  is 0-rectifiable.

*Proof of Part 2:* Let  $\mathbf{x}$  be  $M$ -rectifiable on  $\mathbb{R}^M$ , i.e.,  $\mu\mathbf{x}^{-1} \ll \mathcal{H}^M|_{\mathcal{E}}$  for an  $M$ -rectifiable set  $\mathcal{E}$ . Because  $\mathcal{H}^M$  is equal to the Lebesgue measure  $\mathcal{L}^M$  [18, Th. 2.10.35], we obtain  $\mu\mathbf{x}^{-1} \ll \mathcal{L}^M|_{\mathcal{E}} \ll \mathcal{L}^M$ . Thus, by the Radon-Nikodym theorem, there exists the Radon-Nikodym derivative  $f_{\mathbf{x}} = \frac{d\mu\mathbf{x}^{-1}}{d\mathcal{L}^M}$  satisfying  $\Pr\{\mathbf{x} \in \mathcal{A}\} = \int_{\mathcal{A}} f_{\mathbf{x}}(\mathbf{x}) d\mathcal{L}^M(\mathbf{x})$  for any measurable  $\mathcal{A} \subseteq \mathbb{R}^M$ , i.e.,  $\mathbf{x}$  is a continuous random variable. By (13),  $\theta_{\mathbf{x}}^M = f_{\mathbf{x}} \mathcal{L}^M$ -almost everywhere.

Conversely, let  $\mathbf{x}$  be a continuous random variable on  $\mathbb{R}^M$  with probability density function  $f_{\mathbf{x}}$ . For a measurable set  $\mathcal{A} \subseteq \mathbb{R}^M$  satisfying  $\mathcal{L}^M(\mathcal{A}) = 0$ , we obtain  $\mu\mathbf{x}^{-1}(\mathcal{A}) = \Pr\{\mathbf{x} \in \mathcal{A}\} = \int_{\mathcal{A}} f_{\mathbf{x}}(\mathbf{x}) d\mathcal{L}^M(\mathbf{x}) = 0$ . Thus, we have  $\mu\mathbf{x}^{-1} \ll \mathcal{L}^M$ . Because  $\mathcal{L}^M = \mathcal{H}^M = \mathcal{H}^M|_{\mathbb{R}^M}$ , this is equivalent to  $\mu\mathbf{x}^{-1} \ll \mathcal{H}^M|_{\mathbb{R}^M}$ . Furthermore, by Property 6 in Lemma 4,  $\mathbb{R}^M$  is  $M$ -rectifiable. It then follows from Definition 11 that  $\mathbf{x}$  is an  $M$ -rectifiable random variable.

#### APPENDIX C

##### PROOF OF THEOREM 20

We first note that the set  $\phi(\mathcal{E})$  is  $m$ -rectifiable because  $\mathcal{E}$  is  $m$ -rectifiable and because of Property 3 in Lemma 4. To prove that  $\mathbf{y}$  is  $m$ -rectifiable, we will show that  $\mu\mathbf{y}^{-1} \ll \mathcal{H}^m|_{\phi(\mathcal{E})}$ . For a measurable set  $\mathcal{A} \subseteq \mathbb{R}^M$ , we have

$$\begin{aligned} \mu\mathbf{y}^{-1}(\mathcal{A}) &= \Pr\{\phi(\mathbf{x}) \in \mathcal{A}\} \\ &= \Pr\{\mathbf{x} \in \phi^{-1}(\mathcal{A})\} \\ &\stackrel{(14)}{=} \int_{\phi^{-1}(\mathcal{A})} \theta_{\mathbf{x}}^m(\mathbf{x}) d\mathcal{H}^m|_{\mathcal{E}}(\mathbf{x}) \\ &= \int_{\phi^{-1}(\mathcal{A}) \cap \mathcal{E}} \frac{\theta_{\mathbf{x}}^m(\mathbf{x})}{J_{\phi}^{\mathcal{E}}(\mathbf{x})} J_{\phi}^{\mathcal{E}}(\mathbf{x}) d\mathcal{H}^m(\mathbf{x}) \\ &\stackrel{(a)}{=} \int_{\mathcal{A} \cap \phi(\mathcal{E})} \frac{\theta_{\mathbf{x}}^m(\phi^{-1}(\mathbf{y}))}{J_{\phi}^{\mathcal{E}}(\phi^{-1}(\mathbf{y}))} d\mathcal{H}^m(\mathbf{y}) \\ &= \int_{\mathcal{A}} \frac{\theta_{\mathbf{x}}^m(\phi^{-1}(\mathbf{y}))}{J_{\phi}^{\mathcal{E}}(\phi^{-1}(\mathbf{y}))} d\mathcal{H}^m|_{\phi(\mathcal{E})}(\mathbf{y}). \end{aligned} \quad (106)$$

Here, (a) holds because of the generalized area formula [14, Th. 2.91], and  $\phi^{-1}: \phi(\mathcal{E}) \rightarrow \mathcal{E}$  is well defined because  $\phi$  is one-to-one on  $\mathcal{E}$ . For a measurable set  $\mathcal{A} \subseteq \mathbb{R}^M$  satisfying  $\mathcal{H}^m|_{\phi(\mathcal{E})}(\mathcal{A}) = 0$ , (106) implies  $\mu\mathbf{y}^{-1}(\mathcal{A}) = 0$ , i.e.,  $\mu\mathbf{y}^{-1} \ll \mathcal{H}^m|_{\phi(\mathcal{E})}$ . Thus,  $\mathbf{y}$  is an  $m$ -rectifiable random variable.

By (106),  $\frac{\theta_{\mathbf{x}}^m(\phi^{-1}(\mathbf{y}))}{J_{\phi}^{\mathcal{E}}(\phi^{-1}(\mathbf{y}))}$  equals the Radon-Nikodym derivative  $\frac{d\mu\mathbf{y}^{-1}}{d\mathcal{H}^m|_{\phi(\mathcal{E})}}(\mathbf{y})$ , and thus we obtain

$$\frac{\theta_{\mathbf{x}}^m(\phi^{-1}(\mathbf{y}))}{J_{\phi}^{\mathcal{E}}(\phi^{-1}(\mathbf{y}))} = \frac{d\mu\mathbf{y}^{-1}}{d\mathcal{H}^m|_{\phi(\mathcal{E})}}(\mathbf{y}) \stackrel{(13)}{=} \theta_{\mathbf{y}}^m(\mathbf{y}) \quad (107)$$

for  $\mathcal{H}^m|_{\phi(\mathcal{E})}$ -almost every  $\mathbf{y} \in \mathbb{R}^M$ . We conclude that

$$\begin{aligned} \mathfrak{h}^m(\mathbf{y}) &\stackrel{(21)}{=} - \int_{\phi(\mathcal{E})} \theta_{\mathbf{y}}^m(\mathbf{y}) \log \theta_{\mathbf{y}}^m(\mathbf{y}) d\mathcal{H}^m(\mathbf{y}) \\ &\stackrel{(107)}{=} - \int_{\phi(\mathcal{E})} \frac{\theta_{\mathbf{x}}^m(\phi^{-1}(\mathbf{y}))}{J_{\phi}^{\mathcal{E}}(\phi^{-1}(\mathbf{y}))} \\ &\quad \times \log \left( \frac{\theta_{\mathbf{x}}^m(\phi^{-1}(\mathbf{y}))}{J_{\phi}^{\mathcal{E}}(\phi^{-1}(\mathbf{y}))} \right) d\mathcal{H}^m(\mathbf{y}) \\ &\stackrel{(a)}{=} - \int_{\mathcal{E}} \frac{\theta_{\mathbf{x}}^m(\mathbf{x})}{J_{\phi}^{\mathcal{E}}(\mathbf{x})} \log \left( \frac{\theta_{\mathbf{x}}^m(\mathbf{x})}{J_{\phi}^{\mathcal{E}}(\mathbf{x})} \right) J_{\phi}^{\mathcal{E}}(\mathbf{x}) d\mathcal{H}^m(\mathbf{x}) \\ &= - \int_{\mathcal{E}} \theta_{\mathbf{x}}^m(\mathbf{x}) \log \theta_{\mathbf{x}}^m(\mathbf{x}) d\mathcal{H}^m(\mathbf{x}) \\ &\quad + \int_{\mathcal{E}} \theta_{\mathbf{x}}^m(\mathbf{x}) \log J_{\phi}^{\mathcal{E}}(\mathbf{x}) d\mathcal{H}^m(\mathbf{x}) \\ &\stackrel{(15)}{=} \mathfrak{h}^m(\mathbf{x}) + \mathbb{E}_{\mathbf{x}}[\log J_{\phi}^{\mathcal{E}}(\mathbf{x})] \end{aligned}$$

where (a) holds because of the generalized area formula [14, Th. 2.91].

#### APPENDIX D

##### PROOF OF THEOREM 28

*Proof of Properties 1 and 3:* We first show that for any  $\mu(\mathbf{x}, \mathbf{y})^{-1}$ -measurable set  $\mathcal{A} \subseteq \mathbb{R}^{M_1+M_2}$

$$\begin{aligned} \mu(\mathbf{x}, \mathbf{y})^{-1}(\mathcal{A}) &= \Pr\{(\mathbf{x}, \mathbf{y}) \in \mathcal{A}\} \\ &= \int_{\mathcal{A}} \theta_{\mathbf{x}}^{m_1}(\mathbf{x}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y}) d\mathcal{H}^{m_1+m_2}|_{\mathcal{E}_1 \times \mathcal{E}_2}(\mathbf{x}, \mathbf{y}). \end{aligned} \quad (108)$$

To this end, we first consider the rectangles  $\mathcal{A}_1 \times \mathcal{A}_2$  with  $\mathcal{A}_1 \subseteq \mathbb{R}^{M_1}$   $\mathcal{H}^{m_1}$ -measurable and  $\mathcal{A}_2 \subseteq \mathbb{R}^{M_2}$   $\mathcal{H}^{m_2}$ -measurable. We have

$$\begin{aligned} \Pr\{(\mathbf{x}, \mathbf{y}) \in \mathcal{A}_1 \times \mathcal{A}_2\} &\stackrel{(a)}{=} \Pr\{\mathbf{x} \in \mathcal{A}_1\} \Pr\{\mathbf{y} \in \mathcal{A}_2\} \\ &\stackrel{(14)}{=} \int_{\mathcal{A}_1} \theta_{\mathbf{x}}^{m_1}(\mathbf{x}) d\mathcal{H}^{m_1}|_{\mathcal{E}_1}(\mathbf{x}) \int_{\mathcal{A}_2} \theta_{\mathbf{y}}^{m_2}(\mathbf{y}) d\mathcal{H}^{m_2}|_{\mathcal{E}_2}(\mathbf{y}) \\ &\stackrel{(b)}{=} \int_{\mathcal{A}_1 \times \mathcal{A}_2} \theta_{\mathbf{x}}^{m_1}(\mathbf{x}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y}) d(\mathcal{H}^{m_1}|_{\mathcal{E}_1} \times \mathcal{H}^{m_2}|_{\mathcal{E}_2})(\mathbf{x}, \mathbf{y}) \\ &\stackrel{(c)}{=} \int_{\mathcal{A}_1 \times \mathcal{A}_2} \theta_{\mathbf{x}}^{m_1}(\mathbf{x}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y}) d\mathcal{H}^{m_1+m_2}|_{\mathcal{E}_1 \times \mathcal{E}_2}(\mathbf{x}, \mathbf{y}) \end{aligned} \quad (109)$$

where (a) holds because  $\mathbf{x}$  and  $\mathbf{y}$  are independent, (b) holds by Fubini's theorem, and (c) holds by Lemma 27. Because the rectangles generate the  $\mu(\mathbf{x}, \mathbf{y})^{-1}$ -measurable sets, (109) implies (108). For a  $\mu(\mathbf{x}, \mathbf{y})^{-1}$ -measurable set  $\mathcal{A} \subseteq \mathbb{R}^{M_1+M_2}$  satisfying  $\mathcal{H}^{m_1+m_2}|_{\mathcal{E}_1 \times \mathcal{E}_2}(\mathcal{A}) = 0$ , (108) implies  $\mu(\mathbf{x}, \mathbf{y})^{-1}(\mathcal{A}) = 0$ , i.e.,  $\mu(\mathbf{x}, \mathbf{y})^{-1} \ll \mathcal{H}^{m_1+m_2}|_{\mathcal{E}_1 \times \mathcal{E}_2}$  (note that this is Property 3). Furthermore, since  $\mathbf{x}$  is  $m_1$ -rectifiable and  $\mathbf{y}$  is  $m_2$ -rectifiable, it follows from Lemma 27 that  $\mathcal{E}_1 \times \mathcal{E}_2$  is  $(m_1 + m_2)$ -rectifiable. Hence, according to Definition 11,  $(\mathbf{x}, \mathbf{y})$  is an  $(m_1 + m_2)$ -rectifiable random variable, thus proving Property 1.

*Proof of Property 2:* Again due to (108),

$$\theta_{\mathbf{x}}^{m_1} \theta_{\mathbf{y}}^{m_2} = \frac{d\mu(\mathbf{x}, \mathbf{y})^{-1}}{d\mathcal{H}^{m_1+m_2}|_{\mathcal{E}_1 \times \mathcal{E}_2}} \stackrel{(13)}{=} \theta_{(\mathbf{x}, \mathbf{y})}^{m_1+m_2}.$$



*Proof of Property 4:* We have

$$\begin{aligned}
& \mathfrak{h}^{m_1+m_2}(\mathbf{x}, \mathbf{y}) \\
& \stackrel{(32)}{=} - \int_{\mathbb{R}^{M_1+M_2}} \theta_{(\mathbf{x}, \mathbf{y})}^{m_1+m_2}(\mathbf{x}, \mathbf{y}) \log \theta_{(\mathbf{x}, \mathbf{y})}^{m_1+m_2}(\mathbf{x}, \mathbf{y}) \\
& \quad \times d\mathcal{H}^{m_1+m_2}|_{\mathcal{E}_1 \times \mathcal{E}_2}(\mathbf{x}, \mathbf{y}) \\
& \stackrel{(34)}{=} - \int_{\mathbb{R}^{M_1+M_2}} \theta_{\mathbf{x}}^{m_1}(\mathbf{x}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y}) \log (\theta_{\mathbf{x}}^{m_1}(\mathbf{x}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y})) \\
& \quad \times d\mathcal{H}^{m_1+m_2}|_{\mathcal{E}_1 \times \mathcal{E}_2}(\mathbf{x}, \mathbf{y}) \\
& \stackrel{(a)}{=} - \int_{\mathbb{R}^{M_1+M_2}} \theta_{\mathbf{x}}^{m_1}(\mathbf{x}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y}) \log (\theta_{\mathbf{x}}^{m_1}(\mathbf{x}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y})) \\
& \quad \times d(\mathcal{H}^{m_1}|_{\mathcal{E}_1} \times \mathcal{H}^{m_2}|_{\mathcal{E}_2})(\mathbf{x}, \mathbf{y}) \\
& \stackrel{(b)}{=} - \int_{\mathbb{R}^{M_2}} \int_{\mathbb{R}^{M_1}} \theta_{\mathbf{x}}^{m_1}(\mathbf{x}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y}) \left( \log \theta_{\mathbf{x}}^{m_1}(\mathbf{x}) \right. \\
& \quad \left. + \log \theta_{\mathbf{y}}^{m_2}(\mathbf{y}) \right) d\mathcal{H}^{m_1}|_{\mathcal{E}_1}(\mathbf{x}) d\mathcal{H}^{m_2}|_{\mathcal{E}_2}(\mathbf{y}) \\
& \stackrel{(c)}{=} - \int_{\mathbb{R}^{M_1}} \theta_{\mathbf{x}}^{m_1}(\mathbf{x}) \log \theta_{\mathbf{x}}^{m_1}(\mathbf{x}) d\mathcal{H}^{m_1}|_{\mathcal{E}_1}(\mathbf{x}) \\
& \quad - \int_{\mathbb{R}^{M_2}} \theta_{\mathbf{y}}^{m_2}(\mathbf{y}) \log \theta_{\mathbf{y}}^{m_2}(\mathbf{y}) d\mathcal{H}^{m_2}|_{\mathcal{E}_2}(\mathbf{y}) \\
& \stackrel{(19)}{=} \mathfrak{h}^{m_1}(\mathbf{x}) + \mathfrak{h}^{m_2}(\mathbf{y}).
\end{aligned}$$

Here, (a) holds by Lemma 27, (b) holds by Fubini's theorem, and (c) holds because, by (16),

$$\int_{\mathbb{R}^{M_1}} \theta_{\mathbf{x}}^{m_1}(\mathbf{x}) d\mathcal{H}^{m_1}|_{\mathcal{E}_1}(\mathbf{x}) = \int_{\mathbb{R}^{M_2}} \theta_{\mathbf{y}}^{m_2}(\mathbf{y}) d\mathcal{H}^{m_2}|_{\mathcal{E}_2}(\mathbf{y}) = 1.$$

#### APPENDIX E PROOF OF THEOREM 31

We will use the generalized coarea formula [18, Th. 3.2.22] several times in our proofs. Because the classical version of the generalized coarea formula only holds for sets of finite Hausdorff measure, we first present an adaptation that is suited to our setting.

*Theorem 61:* Let  $\mathcal{E} \subseteq \mathbb{R}^{M_1+M_2}$  be an  $m$ -rectifiable set. Furthermore, let  $\tilde{\mathcal{E}}_2 \triangleq \mathbf{p}_{\mathbf{y}}(\mathcal{E}) \subseteq \mathbb{R}^{M_2}$  be  $m_2$ -rectifiable ( $m_2 \leq M_2$ ,  $m - m_2 \leq M_1$ ),  $\mathcal{H}^{m_2}(\tilde{\mathcal{E}}_2) < \infty$ , and  $J_{\mathbf{p}_{\mathbf{y}}}^{\mathcal{E}} > 0$   $\mathcal{H}^m|_{\mathcal{E}}$ -almost everywhere. Finally, assume that  $g: \mathcal{E} \rightarrow \mathbb{R}$  is  $\mathcal{H}^m$ -measurable and satisfies either of the following properties:

- (i)  $g(\mathbf{x}, \mathbf{y}) \geq 0$   $\mathcal{H}^m$ -almost everywhere;
- (ii)  $\int_{\mathcal{E}} |g(\mathbf{x}, \mathbf{y})| d\mathcal{H}^m(\mathbf{x}, \mathbf{y}) < \infty$ .

Then for all  $\mathcal{H}^{m-m_2}$ -measurable sets  $\mathcal{A}_1 \subseteq \mathbb{R}^{M_1}$  and  $\mathcal{H}^{m_2}$ -measurable sets  $\mathcal{A}_2 \subseteq \mathbb{R}^{M_2}$ ,

$$\begin{aligned}
& \int_{(\mathcal{A}_1 \times \mathcal{A}_2) \cap \mathcal{E}} g(\mathbf{x}, \mathbf{y}) d\mathcal{H}^m(\mathbf{x}, \mathbf{y}) \\
& = \int_{\mathcal{A}_2 \cap \tilde{\mathcal{E}}_2} \int_{\mathcal{A}_1 \cap \mathcal{E}(\mathbf{y})} \frac{g(\mathbf{x}, \mathbf{y})}{J_{\mathbf{p}_{\mathbf{y}}}^{\mathcal{E}}(\mathbf{x}, \mathbf{y})} d\mathcal{H}^{m-m_2}(\mathbf{x}) d\mathcal{H}^{m_2}(\mathbf{y})
\end{aligned} \tag{110}$$

where  $\mathcal{E}(\mathbf{y}) \triangleq \{\mathbf{x} \in \mathbb{R}^{M_1} : (\mathbf{x}, \mathbf{y}) \in \mathcal{E}\}$ . Furthermore, the set  $\mathcal{A}_1 \cap \mathcal{E}(\mathbf{y})$  is  $(m - m_2)$ -rectifiable for  $\mathcal{H}^{m_2}$ -almost every  $\mathbf{y} \in \mathbb{R}^{M_2}$ .

*Proof:* By Property 2 in Lemma 4,  $\mathcal{H}^m|_{\mathcal{E}}$  is  $\sigma$ -finite. Thus, we can partition  $\mathcal{E}$  as  $\mathcal{E} = \bigcup_{i \in \mathbb{N}} \mathcal{F}_i$  with pairwise

disjoint sets  $\mathcal{F}_i$  satisfying  $\mathcal{H}^m(\mathcal{F}_i) < \infty$ . For  $\mathcal{A}_1 \subseteq \mathbb{R}^{M_1}$   $\mathcal{H}^{m_1}$ -measurable and  $\mathcal{A}_2 \subseteq \mathbb{R}^{M_2}$   $\mathcal{H}^{m_2}$ -measurable, we have

$$\begin{aligned}
& \int_{(\mathcal{A}_1 \times \mathcal{A}_2) \cap \mathcal{E}} g(\mathbf{x}, \mathbf{y}) d\mathcal{H}^m(\mathbf{x}, \mathbf{y}) \\
& = \sum_{i \in \mathbb{N}} \int_{(\mathcal{A}_1 \times \mathcal{A}_2) \cap \mathcal{F}_i} g(\mathbf{x}, \mathbf{y}) d\mathcal{H}^m(\mathbf{x}, \mathbf{y}) \\
& \stackrel{(a)}{=} \sum_{i \in \mathbb{N}} \int_{\mathcal{A}_2 \cap \tilde{\mathcal{E}}_2} \int_{(\mathcal{A}_1 \times \mathcal{A}_2) \cap \mathbf{p}_{\mathbf{y}}^{-1}(\{\mathbf{y}\}) \cap \mathcal{F}_i} \frac{g(\mathbf{x}, \mathbf{y}')}{J_{\mathbf{p}_{\mathbf{y}}}^{\mathcal{E}}(\mathbf{x}, \mathbf{y}')} d\mathcal{H}^{m-m_2}(\mathbf{x}, \mathbf{y}') \\
& \quad \times d\mathcal{H}^{m_2}(\mathbf{y})
\end{aligned} \tag{111}$$

where (a) holds by the classical version of the general coarea formula [18, Th. 3.2.22] (note that  $\tilde{\mathcal{E}}_2$  and  $\mathcal{F}_i$  have finite Hausdorff measure) and because  $J_{\mathbf{p}_{\mathbf{y}}}^{\mathcal{E}} > 0$   $\mathcal{H}^m|_{\mathcal{E}}$ -almost everywhere. By either (i) or (ii), we can apply Fubini's theorem in (111) and change the order of integration and summation. We thus obtain

$$\begin{aligned}
& \int_{(\mathcal{A}_1 \times \mathcal{A}_2) \cap \mathcal{E}} g(\mathbf{x}, \mathbf{y}) d\mathcal{H}^m(\mathbf{x}, \mathbf{y}) \\
& = \int_{\mathcal{A}_2 \cap \tilde{\mathcal{E}}_2} \left( \sum_{i \in \mathbb{N}} \int_{(\mathcal{A}_1 \times \mathcal{A}_2) \cap \mathbf{p}_{\mathbf{y}}^{-1}(\{\mathbf{y}\}) \cap \mathcal{F}_i} \frac{g(\mathbf{x}, \mathbf{y}')}{J_{\mathbf{p}_{\mathbf{y}}}^{\mathcal{E}}(\mathbf{x}, \mathbf{y}')} d\mathcal{H}^{m-m_2}(\mathbf{x}, \mathbf{y}') \right) \\
& \quad \times d\mathcal{H}^{m_2}(\mathbf{y}) \\
& = \int_{\mathcal{A}_2 \cap \tilde{\mathcal{E}}_2} \int_{(\mathcal{A}_1 \times \mathcal{A}_2) \cap \mathbf{p}_{\mathbf{y}}^{-1}(\{\mathbf{y}\}) \cap \mathcal{E}} \frac{g(\mathbf{x}, \mathbf{y}')}{J_{\mathbf{p}_{\mathbf{y}}}^{\mathcal{E}}(\mathbf{x}, \mathbf{y}')} d\mathcal{H}^{m-m_2}(\mathbf{x}, \mathbf{y}') d\mathcal{H}^{m_2}(\mathbf{y}) \\
& \stackrel{(a)}{=} \int_{\mathcal{A}_2 \cap \tilde{\mathcal{E}}_2} \int_{(\mathcal{A}_1 \times \mathcal{A}_2) \cap \mathbf{p}_{\mathbf{y}}^{-1}(\{\mathbf{y}\}) \cap \mathcal{E}} \frac{g(\mathbf{x}, \mathbf{y})}{J_{\mathbf{p}_{\mathbf{y}}}^{\mathcal{E}}(\mathbf{x}, \mathbf{y})} d\mathcal{H}^{m-m_2}(\mathbf{x}, \mathbf{y}') d\mathcal{H}^{m_2}(\mathbf{y}) \\
& \stackrel{(b)}{=} \int_{\mathcal{A}_2 \cap \tilde{\mathcal{E}}_2} \int_{\mathcal{A}_1 \cap \mathcal{E}(\mathbf{y})} \frac{g(\mathbf{x}, \mathbf{y})}{J_{\mathbf{p}_{\mathbf{y}}}^{\mathcal{E}}(\mathbf{x}, \mathbf{y})} d\mathcal{H}^{m-m_2}(\mathbf{x}) d\mathcal{H}^{m_2}(\mathbf{y})
\end{aligned}$$

where (a) holds because  $\mathbf{y}' = \mathbf{y}$  for all  $(\mathbf{x}, \mathbf{y}') \in \mathbf{p}_{\mathbf{y}}^{-1}(\{\mathbf{y}\})$ , and (b) holds because the Hausdorff measure does not depend on the ambient space [14, Remark 2.48], i.e., integration with respect to  $\mathcal{H}^{m-m_2}$  on the affine subspace  $\mathbf{p}_{\mathbf{y}}^{-1}(\{\mathbf{y}\}) \subseteq \mathbb{R}^{M_1+M_2}$  and on  $\mathbb{R}^{M_1}$  is identical. Thus, we have shown (110).

We now prove the second part of Theorem 61. By [18, Th. 3.2.22], the sets  $\mathbf{p}_{\mathbf{y}}^{-1}(\{\mathbf{y}\}) \cap \mathcal{F}_i$  are  $(m - m_2)$ -rectifiable for  $\mathcal{H}^{m_2}$ -almost every  $\mathbf{y} \in \mathbb{R}^{M_2}$ . By Property 5 in Lemma 4, the same holds for their union  $\bigcup_{i \in \mathbb{N}} \mathbf{p}_{\mathbf{y}}^{-1}(\{\mathbf{y}\}) \cap \mathcal{F}_i = \mathbf{p}_{\mathbf{y}}^{-1}(\{\mathbf{y}\}) \cap \mathcal{E}$ . The Lipschitz mapping  $\mathbf{p}_{\mathbf{x}}: \mathbb{R}^{M_1+M_2} \rightarrow \mathbb{R}^{M_1}$ ,  $\mathbf{p}_{\mathbf{x}}(\mathbf{x}, \mathbf{y}) = \mathbf{x}$  satisfies

$$\begin{aligned}
& \mathbf{p}_{\mathbf{x}}(\mathbf{p}_{\mathbf{y}}^{-1}(\{\mathbf{y}\}) \cap \mathcal{E}) \\
& = \{\mathbf{x} \in \mathbb{R}^{M_1} : \exists \mathbf{y}' \in \mathbb{R}^{M_2} \text{ with } (\mathbf{x}, \mathbf{y}') \in \mathbf{p}_{\mathbf{y}}^{-1}(\{\mathbf{y}\}) \cap \mathcal{E}\} \\
& = \{\mathbf{x} \in \mathbb{R}^{M_1} : (\mathbf{x}, \mathbf{y}) \in \mathcal{E}\} \\
& = \mathcal{E}(\mathbf{y}).
\end{aligned}$$

Thus,  $\mathcal{E}(\mathbf{y})$  is obtained via a Lipschitz mapping from the set  $\mathbf{p}_{\mathbf{y}}^{-1}(\{\mathbf{y}\}) \cap \mathcal{E}$ , which is  $(m - m_2)$ -rectifiable for  $\mathcal{H}^{m_2}$ -almost every  $\mathbf{y} \in \mathbb{R}^{M_2}$ . Therefore, by Property 3 in Lemma 4,  $\mathcal{E}(\mathbf{y})$

is again  $(m - m_2)$ -rectifiable<sup>19</sup> for  $\mathcal{H}^{m_2}$ -almost every  $\mathbf{y} \in \mathbb{R}^{M_2}$ . Finally, by Property 1 in Lemma 4, the same is true for  $\mathcal{A}_1 \cap \mathcal{E}(\mathbf{y})$ . ■

We now proceed to the proof of Theorem 31.

*Proof of Property 1:* We have for any  $\mathcal{H}^{m_2}$ -measurable set  $\mathcal{A} \subseteq \mathbb{R}^{M_2}$

$$\begin{aligned} & \mu \mathbf{y}^{-1}(\mathcal{A}) \\ &= \Pr\{\mathbf{y} \in \mathcal{A}\} \\ &= \Pr\{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{M_1} \times \mathcal{A}\} \\ &\stackrel{(14)}{=} \int_{(\mathbb{R}^{M_1} \times \mathcal{A}) \cap \mathcal{E}} \theta_{(\mathbf{x}, \mathbf{y})}^m(\mathbf{x}, \mathbf{y}) \, d\mathcal{H}^m(\mathbf{x}, \mathbf{y}) \\ &\stackrel{(a)}{=} \int_{\mathcal{A} \cap \tilde{\mathcal{E}}_2} \int_{\mathcal{E}(\mathbf{y})} \frac{\theta_{(\mathbf{x}, \mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{J_{\mathbf{p}_y}^\mathcal{E}(\mathbf{x}, \mathbf{y})} \, d\mathcal{H}^{m-m_2}(\mathbf{x}) \, d\mathcal{H}^{m_2}(\mathbf{y}) \\ &= \int_{\mathcal{A}} \int_{\mathcal{E}(\mathbf{y})} \frac{\theta_{(\mathbf{x}, \mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{J_{\mathbf{p}_y}^\mathcal{E}(\mathbf{x}, \mathbf{y})} \, d\mathcal{H}^{m-m_2}(\mathbf{x}) \, d\mathcal{H}^{m_2}|_{\tilde{\mathcal{E}}_2}(\mathbf{y}) \quad (112) \end{aligned}$$

where in (a) we used (110) for  $g(\mathbf{x}, \mathbf{y}) = \theta_{(\mathbf{x}, \mathbf{y})}^m(\mathbf{x}, \mathbf{y}) \geq 0$ . For an  $\mathcal{H}^{m_2}$ -measurable set  $\mathcal{A}$  satisfying  $\mathcal{H}^{m_2}|_{\tilde{\mathcal{E}}_2}(\mathcal{A}) = 0$ , (112) implies  $\mu \mathbf{y}^{-1}(\mathcal{A}) = 0$ , i.e.,  $\mu \mathbf{y}^{-1} \ll \mathcal{H}^{m_2}|_{\tilde{\mathcal{E}}_2}$ . Thus, according to Definition 11,  $\mathbf{y}$  is  $m_2$ -rectifiable.

*Proof of Property 2:* Because  $\mu \mathbf{y}^{-1} \ll \mathcal{H}^{m_2}|_{\tilde{\mathcal{E}}_2}$ , it follows from Property 4 in Corollary 12 that there exists a support  $\mathcal{E}_2 \subseteq \tilde{\mathcal{E}}_2$  of the random variable  $\mathbf{y}$ .

*Proof of Property 3:* From (112), we see that

$$\int_{\mathcal{E}(\mathbf{y})} \frac{\theta_{(\mathbf{x}, \mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{J_{\mathbf{p}_y}^\mathcal{E}(\mathbf{x}, \mathbf{y})} \, d\mathcal{H}^{m-m_2}(\mathbf{x}) = \frac{d\mu \mathbf{y}^{-1}}{d\mathcal{H}^{m_2}|_{\tilde{\mathcal{E}}_2}}(\mathbf{y}) = \theta_{\mathbf{y}}^{m_2}(\mathbf{y})$$

where the latter equation holds because  $\mu \mathbf{y}^{-1} \ll \mathcal{H}^{m_2}|_{\tilde{\mathcal{E}}_2}$ . This implies (39).

*Proof of Property 4:* Using (39) in (21) and proceeding similarly to (112), we obtain

$$\begin{aligned} & \mathfrak{h}^{m_2}(\mathbf{y}) \\ &= - \int_{\tilde{\mathcal{E}}_2} \left( \int_{\mathcal{E}(\mathbf{y})} \frac{\theta_{(\mathbf{x}, \mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{J_{\mathbf{p}_y}^\mathcal{E}(\mathbf{x}, \mathbf{y})} \, d\mathcal{H}^{m-m_2}(\mathbf{x}) \right) \\ & \quad \times \log \left( \int_{\mathcal{E}(\mathbf{y})} \frac{\theta_{(\mathbf{x}, \mathbf{y})}^m(\tilde{\mathbf{x}}, \mathbf{y})}{J_{\mathbf{p}_y}^\mathcal{E}(\tilde{\mathbf{x}}, \mathbf{y})} \, d\mathcal{H}^{m-m_2}(\tilde{\mathbf{x}}) \right) \, d\mathcal{H}^{m_2}(\mathbf{y}) \\ &= - \int_{\mathcal{E}} \theta_{(\mathbf{x}, \mathbf{y})}^m(\mathbf{x}, \mathbf{y}) \\ & \quad \times \log \left( \int_{\mathcal{E}(\mathbf{y})} \frac{\theta_{(\mathbf{x}, \mathbf{y})}^m(\tilde{\mathbf{x}}, \mathbf{y})}{J_{\mathbf{p}_y}^\mathcal{E}(\tilde{\mathbf{x}}, \mathbf{y})} \, d\mathcal{H}^{m-m_2}(\tilde{\mathbf{x}}) \right) \, d\mathcal{H}^m(\mathbf{x}, \mathbf{y}). \end{aligned}$$

Thus, (40) holds.

## APPENDIX F PROOF OF THEOREM 35

*Proof of Property 1:* By Theorem 31, the random variable  $\mathbf{y}$  is  $m_2$ -rectifiable with Hausdorff density  $\theta_{\mathbf{y}}^{m_2}$  (given by (39))

<sup>19</sup>Note that  $\mathcal{E}(\mathbf{y})$  is  $\mathcal{H}^{m-m_2}$ -measurable because  $\mathbf{p}_y^{-1}(\{\mathbf{y}\}) \cap \mathcal{E}$  is  $\mathcal{H}^{m-m_2}$ -measurable and the Hausdorff measure does not depend on the ambient space [14, Remark 2.48].

and some support  $\mathcal{E}_2 \subseteq \tilde{\mathcal{E}}_2$ . Let  $\mathcal{A}_1 \subseteq \mathbb{R}^{M_1}$  and  $\mathcal{A}_2 \subseteq \mathbb{R}^{M_2}$  be  $\mathcal{H}^{m_1}$ -measurable and  $\mathcal{H}^{m_2}$ -measurable, respectively. Then

$$\begin{aligned} & \Pr\{(\mathbf{x}, \mathbf{y}) \in \mathcal{A}_1 \times \mathcal{A}_2\} \\ &\stackrel{(44)}{=} \int_{\mathcal{A}_2} \Pr\{\mathbf{x} \in \mathcal{A}_1 \mid \mathbf{y} = \mathbf{y}\} \, d\mu \mathbf{y}^{-1}(\mathbf{y}) \\ &\stackrel{(13)}{=} \int_{\mathcal{A}_2} \Pr\{\mathbf{x} \in \mathcal{A}_1 \mid \mathbf{y} = \mathbf{y}\} \theta_{\mathbf{y}}^{m_2}(\mathbf{y}) \, d\mathcal{H}^{m_2}|_{\mathcal{E}_2}(\mathbf{y}) \\ &\stackrel{(a)}{=} \int_{\mathcal{A}_2} \Pr\{\mathbf{x} \in \mathcal{A}_1 \mid \mathbf{y} = \mathbf{y}\} \theta_{\mathbf{y}}^{m_2}(\mathbf{y}) \, d\mathcal{H}^{m_2}|_{\tilde{\mathcal{E}}_2}(\mathbf{y}) \quad (113) \end{aligned}$$

where (a) holds because we can choose  $\theta_{\mathbf{y}}^{m_2}(\mathbf{y}) = 0$  for  $\mathbf{y} \in \mathcal{E}_2^c$ . On the other hand, we have

$$\begin{aligned} & \Pr\{(\mathbf{x}, \mathbf{y}) \in \mathcal{A}_1 \times \mathcal{A}_2\} \\ &\stackrel{(14)}{=} \int_{(\mathcal{A}_1 \times \mathcal{A}_2) \cap \mathcal{E}} \theta_{(\mathbf{x}, \mathbf{y})}^m(\mathbf{x}, \mathbf{y}) \, d\mathcal{H}^m(\mathbf{x}, \mathbf{y}) \\ &\stackrel{(a)}{=} \int_{\mathcal{A}_2 \cap \tilde{\mathcal{E}}_2} \int_{\mathcal{A}_1 \cap \mathcal{E}(\mathbf{y})} \frac{\theta_{(\mathbf{x}, \mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{J_{\mathbf{p}_y}^\mathcal{E}(\mathbf{x}, \mathbf{y})} \, d\mathcal{H}^{m-m_2}(\mathbf{x}) \, d\mathcal{H}^{m_2}(\mathbf{y}) \\ &= \int_{\mathcal{A}_2} \int_{\mathcal{A}_1 \cap \mathcal{E}(\mathbf{y})} \frac{\theta_{(\mathbf{x}, \mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{J_{\mathbf{p}_y}^\mathcal{E}(\mathbf{x}, \mathbf{y})} \, d\mathcal{H}^{m-m_2}(\mathbf{x}) \, d\mathcal{H}^{m_2}|_{\tilde{\mathcal{E}}_2}(\mathbf{y}) \quad (114) \end{aligned}$$

where in (a) we used (110) for  $g(\mathbf{x}, \mathbf{y}) = \theta_{(\mathbf{x}, \mathbf{y})}^m(\mathbf{x}, \mathbf{y}) \geq 0$ . Combining (113) and (114), we obtain that for  $\mathcal{H}^{m_2}|_{\tilde{\mathcal{E}}_2}$ -almost every  $\mathbf{y}$  and every  $\mathcal{H}^{m_1}$ -measurable set  $\mathcal{A}_1 \subseteq \mathbb{R}^{M_1}$

$$\begin{aligned} & \Pr\{\mathbf{x} \in \mathcal{A}_1 \mid \mathbf{y} = \mathbf{y}\} \theta_{\mathbf{y}}^{m_2}(\mathbf{y}) \\ &= \int_{\mathcal{A}_1 \cap \mathcal{E}(\mathbf{y})} \frac{\theta_{(\mathbf{x}, \mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{J_{\mathbf{p}_y}^\mathcal{E}(\mathbf{x}, \mathbf{y})} \, d\mathcal{H}^{m-m_2}(\mathbf{x}). \quad (115) \end{aligned}$$

Because (115) holds for  $\mathcal{H}^{m_2}|_{\tilde{\mathcal{E}}_2}$ -almost every  $\mathbf{y}$  and  $\mathcal{E}_2 \subseteq \tilde{\mathcal{E}}_2$ , (115) also holds for  $\mathcal{H}^{m_2}|_{\mathcal{E}_2}$ -almost every  $\mathbf{y}$ . Furthermore, because  $\mathcal{E}_2$  is a support of  $\mathbf{y}$ , we have  $\theta_{\mathbf{y}}^{m_2}(\mathbf{y}) > 0$   $\mathcal{H}^{m_2}|_{\mathcal{E}_2}$ -almost everywhere. Thus, we obtain for  $\mathcal{H}^{m_2}|_{\mathcal{E}_2}$ -almost every  $\mathbf{y}$  and every  $\mathcal{H}^{m_1}$ -measurable set  $\mathcal{A}_1 \subseteq \mathbb{R}^{M_1}$

$$\begin{aligned} & \Pr\{\mathbf{x} \in \mathcal{A}_1 \mid \mathbf{y} = \mathbf{y}\} \\ &= \int_{\mathcal{A}_1 \cap \mathcal{E}(\mathbf{y})} \frac{\theta_{(\mathbf{x}, \mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{J_{\mathbf{p}_y}^\mathcal{E}(\mathbf{x}, \mathbf{y}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y})} \, d\mathcal{H}^{m-m_2}(\mathbf{x}) \\ &= \int_{\mathcal{A}_1} \frac{\theta_{(\mathbf{x}, \mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{J_{\mathbf{p}_y}^\mathcal{E}(\mathbf{x}, \mathbf{y}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y})} \, d\mathcal{H}^{m-m_2}|_{\mathcal{E}(\mathbf{y})}(\mathbf{x}). \quad (116) \end{aligned}$$

Therefore,  $\Pr\{\mathbf{x} \in \cdot \mid \mathbf{y} = \mathbf{y}\} \ll \mathcal{H}^{m-m_2}|_{\mathcal{E}(\mathbf{y})}$ . By Theorem 61, the set  $\mathcal{E}(\mathbf{y})$  is  $(m - m_2)$ -rectifiable for  $\mathcal{H}^{m_2}$ -almost every  $\mathbf{y}$ . Hence, according to Definition 6,  $\Pr\{\mathbf{x} \in \cdot \mid \mathbf{y} = \mathbf{y}\}$  is  $(m - m_2)$ -rectifiable for  $\mathcal{H}^{m_2}|_{\mathcal{E}_2}$ -almost every  $\mathbf{y}$ .

*Proof of Property 2:* By (116), we have  $\frac{d\Pr\{\mathbf{x} \in \cdot \mid \mathbf{y} = \mathbf{y}\}}{d\mathcal{H}^{m-m_2}|_{\mathcal{E}(\mathbf{y})}}(\mathbf{x}) = \frac{\theta_{(\mathbf{x}, \mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{J_{\mathbf{p}_y}^\mathcal{E}(\mathbf{x}, \mathbf{y}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y})}$  for  $\mathcal{H}^{m_2}|_{\mathcal{E}_2}$ -almost every  $\mathbf{y}$ . Thus, (10) implies (45).

APPENDIX G  
PROOF OF THEOREM 39

Starting from (47), we have

$$\begin{aligned}
& \mathfrak{h}^{m-m_2}(\mathbf{x} | \mathbf{y}) \\
&= - \int_{\mathcal{E}_2} \theta_{\mathbf{y}}^{m_2}(\mathbf{y}) \int_{\mathcal{E}(\mathbf{y})} \theta_{\text{Pr}\{\mathbf{x} \in \cdot | \mathbf{y}=\mathbf{y}\}}^{m-m_2}(\mathbf{x}) \\
&\quad \times \log \theta_{\text{Pr}\{\mathbf{x} \in \cdot | \mathbf{y}=\mathbf{y}\}}^{m-m_2}(\mathbf{x}) d\mathcal{H}^{m-m_2}(\mathbf{x}) d\mathcal{H}^{m_2}(\mathbf{y}) \\
&\stackrel{(45)}{=} - \int_{\mathcal{E}_2} \theta_{\mathbf{y}}^{m_2}(\mathbf{y}) \int_{\mathcal{E}(\mathbf{y})} \frac{\theta_{(\mathbf{x},\mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{J_{\mathbf{p}_y}^{\mathcal{E}}(\mathbf{x}, \mathbf{y}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y})} \\
&\quad \times \log \left( \frac{\theta_{(\mathbf{x},\mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{J_{\mathbf{p}_y}^{\mathcal{E}}(\mathbf{x}, \mathbf{y}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y})} \right) d\mathcal{H}^{m-m_2}(\mathbf{x}) d\mathcal{H}^{m_2}(\mathbf{y}) \\
&= - \int_{\mathcal{E}_2} \int_{\mathcal{E}(\mathbf{y})} \frac{\theta_{(\mathbf{x},\mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{J_{\mathbf{p}_y}^{\mathcal{E}}(\mathbf{x}, \mathbf{y})} \\
&\quad \times \log \left( \frac{\theta_{(\mathbf{x},\mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{J_{\mathbf{p}_y}^{\mathcal{E}}(\mathbf{x}, \mathbf{y}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y})} \right) d\mathcal{H}^{m-m_2}(\mathbf{x}) d\mathcal{H}^{m_2}(\mathbf{y}) \\
&\stackrel{(a)}{=} - \int_{\mathcal{E}} \theta_{(\mathbf{x},\mathbf{y})}^m(\mathbf{x}, \mathbf{y}) \log \left( \frac{\theta_{(\mathbf{x},\mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{J_{\mathbf{p}_y}^{\mathcal{E}}(\mathbf{x}, \mathbf{y}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y})} \right) d\mathcal{H}^m(\mathbf{x}, \mathbf{y}) \\
&\stackrel{(15)}{=} -\mathbb{E}_{(\mathbf{x},\mathbf{y})} \left[ \log \left( \frac{\theta_{(\mathbf{x},\mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{\theta_{\mathbf{y}}^{m_2}(\mathbf{y})} \right) \right] + \mathbb{E}_{(\mathbf{x},\mathbf{y})} [\log J_{\mathbf{p}_y}^{\mathcal{E}}(\mathbf{x}, \mathbf{y})]
\end{aligned}$$

where in (a) we used (110) with  $\mathcal{A}_1 = \mathbb{R}^{M_1}$ ,  $\mathcal{A}_2 = \mathbb{R}^{M_2}$ , and  $g(\mathbf{x}, \mathbf{y}) = \theta_{(\mathbf{x},\mathbf{y})}^m(\mathbf{x}, \mathbf{y}) \log \left( \frac{\theta_{(\mathbf{x},\mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{J_{\mathbf{p}_y}^{\mathcal{E}}(\mathbf{x}, \mathbf{y}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y})} \right)$ . (Here,  $g(\mathbf{x}, \mathbf{y})$  is  $\mathcal{H}^m|_{\mathcal{E}}$ -integrable by our assumption in Theorem 39 that the right-hand side of (48) exists and is finite, i.e., Condition (ii) in Theorem 61 is satisfied.) Thus, (48) holds.

APPENDIX H  
PROOF OF THEOREM 46

We first note that the product measure  $\mu_{\mathbf{x}^{-1}} \times \mu_{\mathbf{y}^{-1}}$  can be interpreted as the joint measure induced by the independent random variables  $\tilde{\mathbf{x}}$  and  $\tilde{\mathbf{y}}$ , where  $\tilde{\mathbf{x}}$  has the same distribution as  $\mathbf{x}$  and  $\tilde{\mathbf{y}}$  has the same distribution as  $\mathbf{y}$ . Because  $\mathbf{x}$  is  $m_1$ -rectifiable and  $\mathbf{y}$  is  $m_2$ -rectifiable, the same holds for  $\tilde{\mathbf{x}}$  and  $\tilde{\mathbf{y}}$ , respectively. Furthermore, the Hausdorff densities satisfy  $\theta_{\tilde{\mathbf{x}}}^{m_1}(\mathbf{x}) = \theta_{\mathbf{x}}^{m_1}(\mathbf{x})$  and  $\theta_{\tilde{\mathbf{y}}}^{m_2}(\mathbf{y}) = \theta_{\mathbf{y}}^{m_2}(\mathbf{y})$ . By Properties 1–3 in Theorem 28, the joint random variable  $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$  is  $(m_1 + m_2)$ -rectifiable with  $(m_1 + m_2)$ -dimensional Hausdorff density

$$\theta_{(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})}^{m_1+m_2}(\mathbf{x}, \mathbf{y}) = \theta_{\tilde{\mathbf{x}}}^{m_1}(\mathbf{x}) \theta_{\tilde{\mathbf{y}}}^{m_2}(\mathbf{y}) = \theta_{\mathbf{x}}^{m_1}(\mathbf{x}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y}) \quad (117)$$

and  $\mu(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})^{-1} \ll \mathcal{H}^{m_1+m_2}|_{\mathcal{E}_1 \times \mathcal{E}_2}$ . The rectifiability of  $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$  with  $\mu(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})^{-1} \ll \mathcal{H}^{m_1+m_2}|_{\mathcal{E}_1 \times \mathcal{E}_2}$  implies that the measure  $\mu_{\mathbf{x}^{-1}} \times \mu_{\mathbf{y}^{-1}}$  is  $(m_1 + m_2)$ -rectifiable and

$$\mu_{\mathbf{x}^{-1}} \times \mu_{\mathbf{y}^{-1}} \ll \mathcal{H}^{m_1+m_2}|_{\mathcal{E}_1 \times \mathcal{E}_2}. \quad (118)$$

*Proof of Part 1 (case  $m = m_1 + m_2$ ):* For any  $\mathcal{H}^m$ -

measurable set  $\mathcal{A} \subseteq \mathbb{R}^{M_1+M_2}$ , we have

$$\begin{aligned}
& \mu(\mathbf{x}, \mathbf{y})^{-1}(\mathcal{A}) \\
&\stackrel{(14)}{=} \int_{\mathcal{A}} \theta_{(\mathbf{x},\mathbf{y})}^m(\mathbf{x}, \mathbf{y}) d\mathcal{H}^m|_{\mathcal{E}}(\mathbf{x}, \mathbf{y}) \\
&\stackrel{(a)}{=} \int_{\mathcal{A}} \theta_{(\mathbf{x},\mathbf{y})}^m(\mathbf{x}, \mathbf{y}) d\mathcal{H}^m|_{\mathcal{E}_1 \times \mathcal{E}_2}(\mathbf{x}, \mathbf{y}) \\
&\stackrel{(b)}{=} \int_{\mathcal{A}} \frac{\theta_{(\mathbf{x},\mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{\theta_{\mathbf{x}}^{m_1}(\mathbf{x}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y})} \theta_{\mathbf{x}}^{m_1}(\mathbf{x}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y}) d\mathcal{H}^m|_{\mathcal{E}_1 \times \mathcal{E}_2}(\mathbf{x}, \mathbf{y}) \\
&\stackrel{(117)}{=} \int_{\mathcal{A}} \frac{\theta_{(\mathbf{x},\mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{\theta_{\tilde{\mathbf{x}}}^{m_1}(\mathbf{x}) \theta_{\tilde{\mathbf{y}}}^{m_2}(\mathbf{y})} \theta_{(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})}^m(\mathbf{x}, \mathbf{y}) d\mathcal{H}^m|_{\mathcal{E}_1 \times \mathcal{E}_2}(\mathbf{x}, \mathbf{y}) \\
&\stackrel{(c)}{=} \int_{\mathcal{A}} \frac{\theta_{(\mathbf{x},\mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{\theta_{\mathbf{x}}^{m_1}(\mathbf{x}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y})} d(\mu_{\mathbf{x}^{-1}} \times \mu_{\mathbf{y}^{-1}})(\mathbf{x}, \mathbf{y}). \quad (119)
\end{aligned}$$

Here, (a) holds because  $\mathcal{E} \subseteq \mathcal{E}_1 \times \mathcal{E}_2$  and because we can choose  $\theta_{(\mathbf{x},\mathbf{y})}^m(\mathbf{x}, \mathbf{y}) = 0$  on  $\mathcal{E}^c$ , (b) holds because  $\theta_{\mathbf{x}}^{m_1}(\mathbf{x}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y}) > 0$   $\mathcal{H}^m$ -almost everywhere on  $\mathcal{E}_1 \times \mathcal{E}_2$ , and (c) holds because, by (13),  $\theta_{(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})}^m = \frac{d(\mu_{\mathbf{x}^{-1}} \times \mu_{\mathbf{y}^{-1}})}{d\mathcal{H}^m|_{\mathcal{E}_1 \times \mathcal{E}_2}} \mathcal{H}^m|_{\mathcal{E}_1 \times \mathcal{E}_2}$ -almost everywhere. By (119), we obtain that  $\mu(\mathbf{x}, \mathbf{y})^{-1} \ll \mu_{\mathbf{x}^{-1}} \times \mu_{\mathbf{y}^{-1}}$  with Radon-Nikodym derivative

$$\frac{d\mu(\mathbf{x}, \mathbf{y})^{-1}}{d(\mu_{\mathbf{x}^{-1}} \times \mu_{\mathbf{y}^{-1}})}(\mathbf{x}, \mathbf{y}) = \frac{\theta_{(\mathbf{x},\mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{\theta_{\mathbf{x}}^{m_1}(\mathbf{x}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y})}. \quad (120)$$

Inserting (120) into (60) yields

$$\begin{aligned}
I(\mathbf{x}; \mathbf{y}) &= \int_{\mathbb{R}^{M_1+M_2}} \log \left( \frac{\theta_{(\mathbf{x},\mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{\theta_{\mathbf{x}}^{m_1}(\mathbf{x}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y})} \right) d\mu(\mathbf{x}, \mathbf{y})^{-1}(\mathbf{x}, \mathbf{y}) \\
&\stackrel{(13)}{=} \int_{\mathcal{E}} \theta_{(\mathbf{x},\mathbf{y})}^m(\mathbf{x}, \mathbf{y}) \log \left( \frac{\theta_{(\mathbf{x},\mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{\theta_{\mathbf{x}}^{m_1}(\mathbf{x}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y})} \right) d\mathcal{H}^m(\mathbf{x}, \mathbf{y}) \quad (121)
\end{aligned}$$

which is (62). Furthermore, we can rewrite (121) as

$$\begin{aligned}
I(\mathbf{x}; \mathbf{y}) &\stackrel{(15)}{=} \mathbb{E}_{(\mathbf{x},\mathbf{y})} \left[ \log \left( \frac{\theta_{(\mathbf{x},\mathbf{y})}^m(\mathbf{x}, \mathbf{y})}{\theta_{\mathbf{x}}^{m_1}(\mathbf{x}) \theta_{\mathbf{y}}^{m_2}(\mathbf{y})} \right) \right] \\
&= \mathbb{E}_{(\mathbf{x},\mathbf{y})} [\log \theta_{(\mathbf{x},\mathbf{y})}^m(\mathbf{x}, \mathbf{y})] - \mathbb{E}_{(\mathbf{x},\mathbf{y})} [\log \theta_{\mathbf{x}}^{m_1}(\mathbf{x})] \\
&\quad - \mathbb{E}_{(\mathbf{x},\mathbf{y})} [\log \theta_{\mathbf{y}}^{m_2}(\mathbf{y})] \\
&\stackrel{(31)}{=} -\mathfrak{h}^m(\mathbf{x}, \mathbf{y}) - \mathbb{E}_{\mathbf{x}} [\log \theta_{\mathbf{x}}^{m_1}(\mathbf{x})] - \mathbb{E}_{\mathbf{y}} [\log \theta_{\mathbf{y}}^{m_2}(\mathbf{y})] \\
&\stackrel{(19)}{=} -\mathfrak{h}^m(\mathbf{x}, \mathbf{y}) + \mathfrak{h}^{m_1}(\mathbf{x}) + \mathfrak{h}^{m_2}(\mathbf{y}) \quad (122)
\end{aligned}$$

which is (63). Finally, we obtain the first expression in (64) by inserting (56) into (122). The second expression in (64) is obtained by symmetry.

*Proof of Part 2 (case  $m < m_1 + m_2$ ):* We first show that  $\mu(\mathbf{x}, \mathbf{y})^{-1} \not\ll \mu_{\mathbf{x}^{-1}} \times \mu_{\mathbf{y}^{-1}}$ . To this end, we show that the assumption  $\mu(\mathbf{x}, \mathbf{y})^{-1} \ll \mu_{\mathbf{x}^{-1}} \times \mu_{\mathbf{y}^{-1}}$  leads to a contradiction. Using (118), we have  $\mu(\mathbf{x}, \mathbf{y})^{-1} \ll \mu_{\mathbf{x}^{-1}} \times \mu_{\mathbf{y}^{-1}} \ll \mathcal{H}^{m_1+m_2}|_{\mathcal{E}_1 \times \mathcal{E}_2}$ . By Property 4 in Lemma 4 and because  $\mathcal{E}$  is an  $m$ -rectifiable set and  $m_1 + m_2 > m$ , we obtain  $\mathcal{H}^{m_1+m_2}(\mathcal{E}) = 0$ . This implies  $\mathcal{H}^{m_1+m_2}|_{\mathcal{E}_1 \times \mathcal{E}_2}(\mathcal{E}) = 0$ . On the other hand, by (16),  $\mu(\mathbf{x}, \mathbf{y})^{-1}(\mathcal{E}) = 1$ . Thus, we have a contradiction to  $\mu(\mathbf{x}, \mathbf{y})^{-1} \ll \mathcal{H}^{m_1+m_2}|_{\mathcal{E}_1 \times \mathcal{E}_2}$ . Hence,  $\mu(\mathbf{x}, \mathbf{y})^{-1} \not\ll \mu_{\mathbf{x}^{-1}} \times \mu_{\mathbf{y}^{-1}}$  and, by (61),  $I(\mathbf{x}; \mathbf{y}) = \infty$ .

APPENDIX I  
PROOF OF LEMMA 52

Let  $\mathfrak{P}$  denote the set of all finite, measurable partitions of  $\mathbb{R}^M$ , i.e., for  $\Omega = \{\mathcal{A}_1, \dots, \mathcal{A}_N\} \in \mathfrak{P}$  the sets  $\mathcal{A}_i$  are mutually disjoint, measurable, and satisfy  $\bigcup_{i=1}^N \mathcal{A}_i = \mathbb{R}^M$ . Using the interpretation of  $\mathfrak{h}^m(\mathbf{x})$  as a generalized entropy with respect to the Hausdorff measure  $\mathcal{H}^m|_{\mathcal{E}}$  (cf. Remark 19), we obtain by [12, eq. (1.8)]

$$\mathfrak{h}^m(\mathbf{x}) = - \sup_{\Omega \in \mathfrak{P}} \sum_{\mathcal{A} \in \Omega} \mu \mathbf{x}^{-1}(\mathcal{A}) \log \left( \frac{\mu \mathbf{x}^{-1}(\mathcal{A})}{\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A})} \right). \quad (123)$$

Because  $\mu \mathbf{x}^{-1}(\mathcal{E}^c) = 0$  and  $\mathcal{H}^m|_{\mathcal{E}}(\mathcal{E}^c) = 0$ , we have for all  $\Omega \in \mathfrak{P}$

$$\begin{aligned} & \sum_{\mathcal{A} \in \Omega} \mu \mathbf{x}^{-1}(\mathcal{A}) \log \left( \frac{\mu \mathbf{x}^{-1}(\mathcal{A})}{\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A})} \right) \\ &= \sum_{\mathcal{A} \in \Omega} \mu \mathbf{x}^{-1}(\mathcal{A} \cap \mathcal{E}) \log \left( \frac{\mu \mathbf{x}^{-1}(\mathcal{A} \cap \mathcal{E})}{\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A} \cap \mathcal{E})} \right) \\ &= \sum_{\mathcal{A}' \in \tilde{\Omega}} \mu \mathbf{x}^{-1}(\mathcal{A}') \log \left( \frac{\mu \mathbf{x}^{-1}(\mathcal{A}')}{\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}')} \right) \end{aligned} \quad (124)$$

where  $\tilde{\Omega} \triangleq \{\mathcal{A} \cap \mathcal{E} : \mathcal{A} \in \Omega\} \in \mathfrak{P}_{m,\infty}^{(\mathcal{E})}$ . Hence, for every  $\Omega \in \mathfrak{P}$  there exists a  $\tilde{\Omega} \in \mathfrak{P}_{m,\infty}^{(\mathcal{E})}$  such that (124) holds. Thus, the supremum in (123) does not change if we replace  $\mathfrak{P}$  by  $\mathfrak{P}_{m,\infty}^{(\mathcal{E})}$ , i.e., we obtain further

$$\begin{aligned} \mathfrak{h}^m(\mathbf{x}) &= - \sup_{\Omega \in \mathfrak{P}_{m,\infty}^{(\mathcal{E})}} \sum_{\mathcal{A} \in \Omega} \mu \mathbf{x}^{-1}(\mathcal{A}) \log \left( \frac{\mu \mathbf{x}^{-1}(\mathcal{A})}{\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A})} \right) \\ &= \inf_{\Omega \in \mathfrak{P}_{m,\infty}^{(\mathcal{E})}} \left( - \sum_{\mathcal{A} \in \Omega} \mu \mathbf{x}^{-1}(\mathcal{A}) \log \left( \frac{\mu \mathbf{x}^{-1}(\mathcal{A})}{\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A})} \right) \right) \\ &= \inf_{\Omega \in \mathfrak{P}_{m,\infty}^{(\mathcal{E})}} \left( - \sum_{\mathcal{A} \in \Omega} \mu \mathbf{x}^{-1}(\mathcal{A}) \log \mu \mathbf{x}^{-1}(\mathcal{A}) \right. \\ &\quad \left. + \sum_{\mathcal{A} \in \Omega} \mu \mathbf{x}^{-1}(\mathcal{A}) \log \mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}) \right) \\ &= \inf_{\Omega \in \mathfrak{P}_{m,\infty}^{(\mathcal{E})}} \left( H([\mathbf{x}]_{\Omega}) + \sum_{\mathcal{A} \in \Omega} \mu \mathbf{x}^{-1}(\mathcal{A}) \log \mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}) \right). \end{aligned} \quad (126)$$

Here, (125) is (68) and (126) is (69).

APPENDIX J  
PROOF OF THEOREM 53

J.1 Proof of Lower Bound (70)

Let  $\Omega \in \mathfrak{P}_{m,\delta}^{(\mathcal{E})}$  be an  $(m, \delta)$ -partition of  $\mathcal{E}$  according to Definition 51, i.e.,  $\Omega = \{\mathcal{A}_1, \dots, \mathcal{A}_N\}$  where  $\bigcup_{i=1}^N \mathcal{A}_i = \mathcal{E}$ ,  $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset$ , and  $\mathcal{H}^m(\mathcal{A}_i) \leq \delta$  for all  $i, j \in \{1, \dots, N\}$ ,  $i \neq j$ . Note that  $\Omega$  also belongs to  $\mathfrak{P}_{m,\infty}^{(\mathcal{E})}$ . Then, starting from

(69), we obtain

$$\begin{aligned} \mathfrak{h}^m(\mathbf{x}) &= \inf_{\Omega' \in \mathfrak{P}_{m,\infty}^{(\mathcal{E})}} \left( H([\mathbf{x}]_{\Omega'}) \right. \\ &\quad \left. + \sum_{\mathcal{A} \in \Omega'} \mu \mathbf{x}^{-1}(\mathcal{A}) \log \mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}) \right) \\ &\leq H([\mathbf{x}]_{\Omega}) + \sum_{i=1}^N \mu \mathbf{x}^{-1}(\mathcal{A}_i) \log \mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_i) \\ &\stackrel{(a)}{\leq} H([\mathbf{x}]_{\Omega}) + \sum_{i=1}^N \mu \mathbf{x}^{-1}(\mathcal{A}_i) \log \delta \\ &\stackrel{(b)}{=} H([\mathbf{x}]_{\Omega}) + \log \delta \end{aligned}$$

where (a) holds because  $\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_i) \leq \delta$  and (b) holds because  $\sum_{i=1}^N \mu \mathbf{x}^{-1}(\mathcal{A}_i) = \mu \mathbf{x}^{-1}(\mathcal{E}) = 1$ . Multiplying by  $\text{ld } e$ , we have equivalently

$$(\mathfrak{h}^m(\mathbf{x}) - \log \delta) \text{ld } e \leq H([\mathbf{x}]_{\Omega}) \text{ld } e. \quad (127)$$

By (67), we have

$$H([\mathbf{x}]_{\Omega}) \text{ld } e \leq L^*([\mathbf{x}]_{\Omega}). \quad (128)$$

Combining (127) and (128), we obtain

$$(\mathfrak{h}^m(\mathbf{x}) - \log \delta) \text{ld } e \leq L^*([\mathbf{x}]_{\Omega})$$

which implies (70).

J.2 Proof of Upper Bound (71)

We first state a preliminary result.

*Lemma 62:* Let  $\mathbf{x}$  be an  $m$ -rectifiable random variable, i.e.,  $\mu \mathbf{x}^{-1} \ll \mathcal{H}^m|_{\mathcal{E}}$  for an  $m$ -rectifiable set  $\mathcal{E} \subseteq \mathbb{R}^M$ , with  $m \in \{1, \dots, M\}$  and  $\mathcal{H}^m(\mathcal{E}) < \infty$ . Furthermore, let  $\Omega = \{\mathcal{A}_1, \dots, \mathcal{A}_N\} \in \mathfrak{P}_{m,\infty}^{(\mathcal{E})}$ , where each  $\mathcal{A}_i$  is constructed as the union of disjoint sets  $\mathcal{A}_{i,1}, \dots, \mathcal{A}_{i,k_i}$ , i.e.,  $\mathcal{A}_i = \bigcup_{j=1}^{k_i} \mathcal{A}_{i,j}$  with  $\mathcal{A}_{i,j_1} \cap \mathcal{A}_{i,j_2} = \emptyset$  for  $j_1 \neq j_2$ . Finally, let  $\tilde{\Omega} \triangleq \{\mathcal{A}_{1,1}, \dots, \mathcal{A}_{1,k_1}, \dots, \mathcal{A}_{N,1}, \dots, \mathcal{A}_{N,k_N}\}$ . Then

$$\begin{aligned} & - \sum_{\mathcal{A} \in \Omega} \mu \mathbf{x}^{-1}(\mathcal{A}) \log \left( \frac{\mu \mathbf{x}^{-1}(\mathcal{A})}{\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A})} \right) \\ & \geq - \sum_{\mathcal{A} \in \tilde{\Omega}} \mu \mathbf{x}^{-1}(\mathcal{A}) \log \left( \frac{\mu \mathbf{x}^{-1}(\mathcal{A})}{\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A})} \right). \end{aligned} \quad (129)$$

*Proof:* The inequality (129) can be written as

$$\begin{aligned} & - \sum_{i=1}^N \mu \mathbf{x}^{-1}(\mathcal{A}_i) \log \left( \frac{\mu \mathbf{x}^{-1}(\mathcal{A}_i)}{\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_i)} \right) \\ & \geq - \sum_{i=1}^N \sum_{j=1}^{k_i} \mu \mathbf{x}^{-1}(\mathcal{A}_{i,j}) \log \left( \frac{\mu \mathbf{x}^{-1}(\mathcal{A}_{i,j})}{\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_{i,j})} \right). \end{aligned}$$

Therefore, it suffices to show that

$$\begin{aligned} & \mu \mathbf{x}^{-1}(\mathcal{A}_i) \log \left( \frac{\mu \mathbf{x}^{-1}(\mathcal{A}_i)}{\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_i)} \right) \\ & \leq \sum_{j=1}^{k_i} \mu \mathbf{x}^{-1}(\mathcal{A}_{i,j}) \log \left( \frac{\mu \mathbf{x}^{-1}(\mathcal{A}_{i,j})}{\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_{i,j})} \right) \end{aligned}$$

for  $i \in \{1, \dots, N\}$ . This latter inequality follows from the log sum inequality [13, Th. 2.7.1]. ■

We now proceed to the proof of (71). By (68), for each  $\varepsilon' > 0$ , there exists a partition  $\Omega = \{\mathcal{A}_1, \dots, \mathcal{A}_N\} \in \mathfrak{P}_{m,\infty}^{(\varepsilon)}$  such that

$$\mathfrak{h}^m(\mathbf{x}) > - \sum_{\mathcal{A} \in \Omega} \mu \mathbf{x}^{-1}(\mathcal{A}) \log \left( \frac{\mu \mathbf{x}^{-1}(\mathcal{A})}{\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A})} \right) - \varepsilon'. \quad (130)$$

Let us choose, in particular,  $\varepsilon' \triangleq \frac{\varepsilon}{2 \ln e}$ . We define

$$\delta_\varepsilon \triangleq (1 - e^{-\varepsilon'}) \min_{\substack{\mathcal{A}_i \in \Omega \\ \mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_i) \neq 0}} \mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_i) > 0. \quad (131)$$

Choosing some  $\delta \in (0, \delta_\varepsilon)$ , we furthermore define

$$J_{i,\delta} \triangleq \frac{\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_i)}{\delta}$$

and

$$M_{i,\delta} \triangleq \begin{cases} \lceil J_{i,\delta} \rceil & \text{if } \mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_i) \neq 0 \\ 1 & \text{if } \mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_i) = 0. \end{cases}$$

Let us partition each set  $\mathcal{A}_i \in \Omega$  into  $M_{i,\delta}$  disjoint subsets  $\mathcal{A}_{i,j}$  of equal Hausdorff measure,<sup>20</sup> i.e.,

$$\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_{i,j}) = \frac{\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_i)}{M_{i,\delta}}$$

and

$$\bigcup_{j=1}^{M_{i,\delta}} \mathcal{A}_{i,j} = \mathcal{A}_i.$$

For  $\mathcal{A}_i \in \Omega$  such that  $\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_i) = 0$ , we have  $M_{i,\delta} = 1$ , and thus this partition degenerates to  $\mathcal{A}_{i,1} = \mathcal{A}_i$ , which implies  $\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_{i,1}) = 0$ . For  $\mathcal{A}_i \in \Omega$  such that  $\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_i) \neq 0$ , we have  $M_{i,\delta} = \lceil J_{i,\delta} \rceil$ , and thus

$$\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_{i,j}) = \frac{\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_i)}{\lceil J_{i,\delta} \rceil} = \frac{J_{i,\delta}}{\lceil J_{i,\delta} \rceil} \delta \leq \delta. \quad (132)$$

In either case we have  $\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_{i,j}) \leq \delta$ .

Let us denote by  $\Omega_\delta$  the partition of  $\mathcal{E}$  containing all the sets  $\mathcal{A}_{i,j}$ . Then  $\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_{i,j}) \leq \delta$  implies  $\Omega_\delta \in \mathfrak{P}_{m,\delta}^{(\varepsilon)}$ . Furthermore, for  $\mathcal{A}_{i,j} \in \Omega_\delta$  satisfying  $\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_{i,j}) \neq 0$ ,

$$\begin{aligned} \mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_{i,j}) &\stackrel{(132)}{=} \frac{J_{i,\delta}}{\lceil J_{i,\delta} \rceil} \delta \\ &= \frac{\lceil J_{i,\delta} \rceil - (\lceil J_{i,\delta} \rceil - J_{i,\delta})}{\lceil J_{i,\delta} \rceil} \delta \\ &= \left( 1 - \frac{\lceil J_{i,\delta} \rceil - J_{i,\delta}}{\lceil J_{i,\delta} \rceil} \right) \delta \\ &\stackrel{(a)}{>} \left( 1 - \frac{1}{\lceil J_{i,\delta} \rceil} \right) \delta \end{aligned} \quad (133)$$

where (a) holds because  $\lceil J_{i,\delta} \rceil - J_{i,\delta} < 1$ . Furthermore, we can bound  $\lceil J_{i,\delta} \rceil$  as (note that  $\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_{i,j}) \neq 0$  implies

<sup>20</sup>Because  $\mathcal{H}^m$  is a nonatomic measure, we can always find subsets of arbitrary but smaller measure (see [29, Sec. 2.5]).

$$\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_i) \neq 0)$$

$$\begin{aligned} \lceil J_{i,\delta} \rceil &\geq J_{i,\delta} = \frac{\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_i)}{\delta} \\ &> \frac{\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_i)}{\delta_\varepsilon} \\ &\stackrel{(131)}{=} \frac{\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_i)}{(1 - e^{-\varepsilon'}) \min_{\substack{i' \in \{1, \dots, N\} \\ \mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_{i'}) \neq 0}} \mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_{i'})} \\ &\geq \frac{1}{1 - e^{-\varepsilon'}}. \end{aligned} \quad (134)$$

Inserting (134) into (133), we obtain for all sets  $\mathcal{A}_{i,j} \in \Omega_\delta$  satisfying  $\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_{i,j}) \neq 0$

$$\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}_{i,j}) > \left( 1 - \frac{1}{\frac{1}{1 - e^{-\varepsilon'}}} \right) \delta = e^{-\varepsilon'} \delta. \quad (135)$$

Combining our results yields

$$\begin{aligned} \mathfrak{h}^m(\mathbf{x}) &\stackrel{(130)}{>} - \sum_{\mathcal{A} \in \Omega} \mu \mathbf{x}^{-1}(\mathcal{A}) \log \left( \frac{\mu \mathbf{x}^{-1}(\mathcal{A})}{\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A})} \right) - \varepsilon' \\ &\stackrel{(129)}{\geq} - \sum_{\mathcal{A} \in \Omega_\delta} \mu \mathbf{x}^{-1}(\mathcal{A}) \log \left( \frac{\mu \mathbf{x}^{-1}(\mathcal{A})}{\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A})} \right) - \varepsilon' \\ &\stackrel{(a)}{\geq} - \sum_{\substack{\mathcal{A} \in \Omega_\delta \\ \mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}) \neq 0}} \mu \mathbf{x}^{-1}(\mathcal{A}) \log \left( \frac{\mu \mathbf{x}^{-1}(\mathcal{A})}{\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A})} \right) - \varepsilon' \\ &\stackrel{(135)}{>} - \sum_{\substack{\mathcal{A} \in \Omega_\delta \\ \mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}) \neq 0}} \mu \mathbf{x}^{-1}(\mathcal{A}) \log \left( \frac{\mu \mathbf{x}^{-1}(\mathcal{A})}{e^{-\varepsilon'} \delta} \right) - \varepsilon' \\ &\stackrel{(b)}{=} - \sum_{\mathcal{A} \in \Omega_\delta} \mu \mathbf{x}^{-1}(\mathcal{A}) \log \left( \frac{\mu \mathbf{x}^{-1}(\mathcal{A})}{e^{-\varepsilon'} \delta} \right) - \varepsilon' \\ &\stackrel{(c)}{=} - \sum_{\mathcal{A} \in \Omega_\delta} \mu \mathbf{x}^{-1}(\mathcal{A}) \log (\mu \mathbf{x}^{-1}(\mathcal{A})) + \log \delta - 2\varepsilon' \\ &= H([\mathbf{x}]_{\Omega_\delta}) + \log \delta - 2\varepsilon' \\ &\stackrel{(d)}{>} \frac{L^*([\mathbf{x}]_{\Omega_\delta}) - 1}{\ln e} + \log \delta - 2\varepsilon' \end{aligned} \quad (136)$$

where (a) and (b) hold because, by  $\mu \mathbf{x}^{-1} \ll \mathcal{H}^m|_{\mathcal{E}}$ ,  $\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}) = 0$  implies  $\mu \mathbf{x}^{-1}(\mathcal{A}) = 0$  and thus the additional restriction  $\mathcal{H}^m|_{\mathcal{E}}(\mathcal{A}) \neq 0$  removes only summands that are zero, (c) holds because  $\Omega_\delta$  is a partition of  $\mathcal{E}$  and thus  $\sum_{\mathcal{A} \in \Omega_\delta} \mu \mathbf{x}^{-1}(\mathcal{A}) = \mu \mathbf{x}^{-1}(\mathcal{E}) = 1$ , and (d) holds by the second inequality in (67). Finally, rewriting (136) gives (recall  $\varepsilon' = \frac{\varepsilon}{2 \ln e}$ )

$$L^*([\mathbf{x}]_{\Omega_\delta}) < \mathfrak{h}^m(\mathbf{x}) \ln e - \log \delta \ln e + 1 + \varepsilon$$

which is (71).

## APPENDIX K

### PROOF OF LEMMA 56

Because  $R_{\text{SLB}}(D, s) = \mathfrak{h}^m(\mathbf{x}) - (sD + \log \gamma(s))$ , where  $\mathfrak{h}^m(\mathbf{x})$  is finite and does not depend on  $s$ , it is sufficient to show  $\lim_{s \rightarrow \infty} (sD + \log \gamma(s)) = \infty$ . For  $\mathbf{y} \in \mathbb{R}^M$ , we define the set of all  $\mathbf{x}$  whose distortion relative to  $\mathbf{y}$  is less than  $D/2$ ,

$$\mathcal{C}(\mathbf{y}) \triangleq \left\{ \mathbf{x} \in \mathbb{R}^M : d(\mathbf{x}, \mathbf{y}) < \frac{D}{2} \right\}.$$

We obtain

$$\begin{aligned}
& sD + \log \gamma(s) \\
& \stackrel{(80)}{=} sD + \log \left( \sup_{\mathbf{y} \in \mathbb{R}^M} \int_{\mathcal{E}} e^{-sd(\mathbf{x}, \mathbf{y})} d\mathcal{H}^m(\mathbf{x}) \right) \\
& \stackrel{(a)}{=} \sup_{\mathbf{y} \in \mathbb{R}^M} \left( sD + \log \left( \int_{\mathcal{E}} e^{-sd(\mathbf{x}, \mathbf{y})} d\mathcal{H}^m(\mathbf{x}) \right) \right) \\
& = \sup_{\mathbf{y} \in \mathbb{R}^M} \log \left( e^{sD} \int_{\mathcal{E}} e^{-sd(\mathbf{x}, \mathbf{y})} d\mathcal{H}^m(\mathbf{x}) \right) \\
& = \sup_{\mathbf{y} \in \mathbb{R}^M} \log \left( \int_{\mathcal{E}} e^{s(D-d(\mathbf{x}, \mathbf{y}))} d\mathcal{H}^m(\mathbf{x}) \right) \\
& \stackrel{(b)}{\geq} \sup_{\mathbf{y} \in \mathbb{R}^M} \log \left( \int_{\mathcal{E} \cap \mathcal{C}(\mathbf{y})} e^{s(D-d(\mathbf{x}, \mathbf{y}))} d\mathcal{H}^m(\mathbf{x}) \right) \\
& \stackrel{(c)}{\geq} \sup_{\mathbf{y} \in \mathbb{R}^M} \log \left( \int_{\mathcal{E} \cap \mathcal{C}(\mathbf{y})} e^{sD/2} d\mathcal{H}^m(\mathbf{x}) \right) \\
& = \sup_{\mathbf{y} \in \mathbb{R}^M} \log (e^{sD/2} \mathcal{H}^m(\mathcal{E} \cap \mathcal{C}(\mathbf{y}))) \\
& = s \frac{D}{2} + \sup_{\mathbf{y} \in \mathbb{R}^M} \log \mathcal{H}^m(\mathcal{E} \cap \mathcal{C}(\mathbf{y})) \quad (137)
\end{aligned}$$

where (a) holds because  $\log$  is a monotonically increasing function, (b) holds because  $e^{s(D-d(\mathbf{x}, \mathbf{y}))} > 0$ , and (c) holds because  $d(\mathbf{x}, \mathbf{y}) < D/2$  for all  $\mathbf{x} \in \mathcal{C}(\mathbf{y})$ . Because  $\mu_{\mathbf{x}^{-1}}(\mathcal{E}) = 1$  (see (16)), the absolute continuity  $\mu_{\mathbf{x}^{-1}} \ll \mathcal{H}^m|_{\mathcal{E}}$  implies  $\mathcal{H}^m(\mathcal{E}) > 0$ . Thus, there exists a  $\bar{\mathbf{y}} \in \mathbb{R}^M$  such that  $\delta \triangleq \mathcal{H}^m(\mathcal{E} \cap \mathcal{C}(\bar{\mathbf{y}})) > 0$ . Clearly, this implies  $\sup_{\mathbf{y} \in \mathbb{R}^M} \log \mathcal{H}^m(\mathcal{E} \cap \mathcal{C}(\mathbf{y})) \geq \log \delta$ , and hence, by (137),

$$sD + \log \gamma(s) \geq s \frac{D}{2} + \log \delta. \quad (138)$$

For fixed but arbitrary  $K > 0$  and all  $s \geq \frac{2(K - \log \delta)}{D}$ , we have  $s \frac{D}{2} + \log \delta \geq K$ , and thus (138) implies

$$sD + \log \gamma(s) \geq K.$$

Since  $K$  can be chosen arbitrarily large, this shows that  $\lim_{s \rightarrow \infty} (sD + \log \gamma(s)) = \infty$ .

#### APPENDIX L PROOF OF THEOREM 60

Consider the source  $\mathbf{x}$  on  $\mathbb{R}^2$  as specified in Theorem 60. The main idea of the proof is to construct a specific source code and calculate its rate and expected distortion. We can then use the source coding theorem [25, Th. 11.4.1] to conclude that the calculated rate is an upper bound on the RD function.

To this end, recall that a  $(k, n)$  source code for a sequence  $\mathbf{x}_{1:k} \in (\mathbb{R}^2)^k$  of  $k$  independent realizations of  $\mathbf{x}$  consists of an encoding function  $f: (\mathbb{R}^2)^k \rightarrow \{1, \dots, n\}$  and a decoding function  $g: \{1, \dots, n\} \rightarrow (\mathbb{R}^2)^k$ . The rate of this code is defined as  $R_{f,g} \triangleq (\log n)/k$  and the expected distortion is given by

$$D_{f,g} = \mathbb{E}_{\mathbf{x}_{1:k}} [\|\mathbf{x}_{1:k} - g(f(\mathbf{x}_{1:k}))\|^2].$$

By the source coding theorem [25, Th. 11.4.1], every  $(k, n)$  code with expected distortion  $D_{f,g}$  must have a rate greater than or equal to  $R(D_{f,g})$ . In particular, this has to hold for

the special case  $k = 1$ . The rate of these  $(1, n)$  codes reduces to  $R_{f,g} = \log n$ , and the expected distortion is given by

$$D_{f,g} = \mathbb{E}_{\mathbf{x}} [\|\mathbf{x} - g(f(\mathbf{x}))\|^2]. \quad (139)$$

Thus, the implication of the source coding theorem is that for a  $(1, n)$  code with expected distortion  $D_{f,g}$ , we have

$$\log n \geq R(D_{f,g}). \quad (140)$$

We directly design the composed function  $q \triangleq g \circ f$ . Because  $\mathbf{x}$  has probability zero outside  $\mathcal{S}_1$ , we only have to define  $q$  on the unit circle. Furthermore, because  $f$  maps  $\mathbf{x}$  to one of at most  $n$  distinct values,  $q = g \circ f$  can also attain at most  $n$  distinct values. We define  $q$  to map each circle segment defined by an angle interval  $[i \frac{2\pi}{n}, (i+1) \frac{2\pi}{n})$ ,  $i \in \{0, \dots, n-1\}$ , onto one associated “center” point, which is not constrained to lie on the unit circle. To this end, we only have to consider the circle segment defined by  $\{\mathbf{x} = (\cos \phi \ \sin \phi)^T : \phi \in [-\pi/n, \pi/n)\}$  since the problem is invariant under rotations. Because of symmetry, we choose the “center” associated with this segment to be some point  $(x_1 \ 0)^T$ , i.e.,  $q(\mathbf{x}) = (x_1 \ 0)^T$  for all  $\mathbf{x} = (\cos \phi \ \sin \phi)^T$  with  $\phi \in [-\pi/n, \pi/n)$ . According to (139), the expected distortion is then obtained as

$$\begin{aligned}
D_q &= \mathbb{E}_{\mathbf{x}} [\|\mathbf{x} - q(\mathbf{x})\|^2] \\
&= \int_0^{2\pi} \frac{1}{2\pi} \left\| \begin{pmatrix} \cos \phi \\ \sin \phi \end{pmatrix} - q \left( \begin{pmatrix} \cos \phi \\ \sin \phi \end{pmatrix} \right) \right\|^2 d\phi \\
&= \frac{n}{2\pi} \int_{-\pi/n}^{\pi/n} \left\| \begin{pmatrix} \cos \phi \\ \sin \phi \end{pmatrix} - \begin{pmatrix} x_1 \\ 0 \end{pmatrix} \right\|^2 d\phi \\
&= \frac{n}{2\pi} \int_{-\pi/n}^{\pi/n} ((\cos \phi - x_1)^2 + \sin^2 \phi) d\phi \\
&= \frac{n}{2\pi} \int_{-\pi/n}^{\pi/n} (1 + x_1^2 - 2x_1 \cos \phi) d\phi \\
&= 1 + x_1^2 - \frac{2nx_1}{\pi} \sin \frac{\pi}{n}. \quad (141)
\end{aligned}$$

Minimizing the expected distortion with respect to  $x_1$  gives the optimum value of  $x_1$  as

$$x_1^* = \frac{n}{\pi} \sin \frac{\pi}{n}. \quad (142)$$

The corresponding quantization function will be denoted by  $q^*$ . Inserting (142) into (141) yields  $\bar{D}_n$  in (98):

$$\begin{aligned}
D_{q^*} &= \mathbb{E}_{\mathbf{x}} [\|\mathbf{x} - q^*(\mathbf{x})\|^2] \\
&= 1 + \left( \frac{n}{\pi} \sin \frac{\pi}{n} \right)^2 - 2 \left( \frac{n}{\pi} \sin \frac{\pi}{n} \right)^2 \\
&= 1 - \left( \frac{n}{\pi} \sin \frac{\pi}{n} \right)^2 \\
&= \bar{D}_n.
\end{aligned}$$

Thus, we found a  $(1, n)$  code with expected distortion  $D_{q^*} = \bar{D}_n$ . Hence, by (140), we have  $\log n \geq R(D_{q^*}) = R(\bar{D}_n)$ , which is (97).

## ACKNOWLEDGMENT

The authors wish to thank the anonymous reviewers for numerous insightful comments and, in particular, for suggesting reference [12].

## REFERENCES

- [1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 623–656, July, Oct. 1948.
- [2] I. Csiszár, "Axiomatic characterizations of information measures," *Entropy*, vol. 10, pp. 261–273, 2008.
- [3] A. Kolmogorov, "On the Shannon theory of information transmission in the case of continuous signals," *IRE Trans. Inf. Theory*, vol. 2, no. 4, pp. 102–108, Dec. 1956.
- [4] R. M. Gray, *Source Coding Theory*. Boston, MA: Kluwer, 1990.
- [5] D. Stotz and H. Bölcskei, "Degrees of freedom in vector interference channels," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 4172–4197, July 2016.
- [6] Y. Wu and S. Verdú, "Rényi information dimension: Fundamental limits of almost lossless analog compression," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3721–3748, Aug. 2010.
- [7] R. Palanki, "Iterative decoding for wireless networks," Ph.D. dissertation, California Inst. of Technol., Pasadena, CA, 2004.
- [8] G. Koliander, E. Riegler, G. Durisi, and F. Hlawatsch, "Degrees of freedom of generic block-fading MIMO channels without a priori channel state information," *IEEE Trans. Inf. Theory*, vol. 60, no. 12, pp. 7760–7781, Dec. 2014.
- [9] A. Rényi, "On the dimension and entropy of probability distributions," *Acta Math. Hung.*, vol. 10, no. 1, pp. 193–215, 1959.
- [10] C. E. Posner and E. R. Rodemich, "Epsilon entropy and data compression," *Ann. Math. Statist.*, vol. 42, no. 6, pp. 2079–2125, 1971.
- [11] C. De Lellis, *Rectifiable Sets, Densities and Tangent Measures*, ser. Zurich Lectures in Advanced Mathematics. Zurich, Switzerland: EMS Publishing House, 2008.
- [12] I. Csiszár, "Generalized entropy and quantization problems," in *Proc. Sixth Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, Prague, Czechoslovakia, Sep. 1971, pp. 159–174.
- [13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY: Wiley, 2006.
- [14] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*. New York, NY: Oxford Univ. Press, 2000.
- [15] T. Kawabata and A. Dembo, "The rate-distortion dimension of sets and measures," *IEEE Trans. Inf. Theory*, vol. 40, no. 5, pp. 1564–1572, Sep. 1994.
- [16] D. Preiss, "Geometry of measures in  $\mathbb{R}^n$ : Distribution, rectifiability, and densities," *Ann. Math.*, vol. 125, no. 3, pp. 537–643, May 1987.
- [17] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.
- [18] H. Federer, *Geometric Measure Theory*. New York, NY: Springer, 1969.
- [19] S. G. Krantz and H. R. Parks, *Geometric Integration Theory*. Basel, Switzerland: Birkhäuser, 2009.
- [20] A. S. Kechris, *Classical Descriptive Set Theory*. New York: Springer, 1995.
- [21] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. Cambridge, UK: Cambridge Univ. Press, 2013.
- [22] H. Uhlig, "On singular Wishart and singular multivariate beta distributions," *Ann. Statist.*, vol. 22, no. 1, pp. 395–405, 1994.
- [23] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY: Springer, 2006.
- [24] R. M. Gray, *Probability, Random Processes, and Ergodic Properties*. New York, NY: Springer, 2010.
- [25] R. M. Gray, *Entropy and Information Theory*. New York, NY: Springer, 1990.
- [26] I. Csiszár, "On an extremum problem of information theory," *Stud. Sci. Math. Hung.*, vol. 9, no. 1, pp. 57–71, 1974.
- [27] R. G. Bartle, *The Elements of Integration and Lebesgue Measure*. New York, NY: Wiley, 1995.
- [28] P. Mattila, *Geometry of Sets and Measures in Euclidean Spaces*. Cambridge, UK: Cambridge Univ. Press, 1995.
- [29] A. Fryszkowski, *Fixed Point Theory for Decomposable Sets*, ser. Topological Fixed Point Theory and Its Applications. Dordrecht, The Netherlands: Kluwer, 2004, vol. 2.